

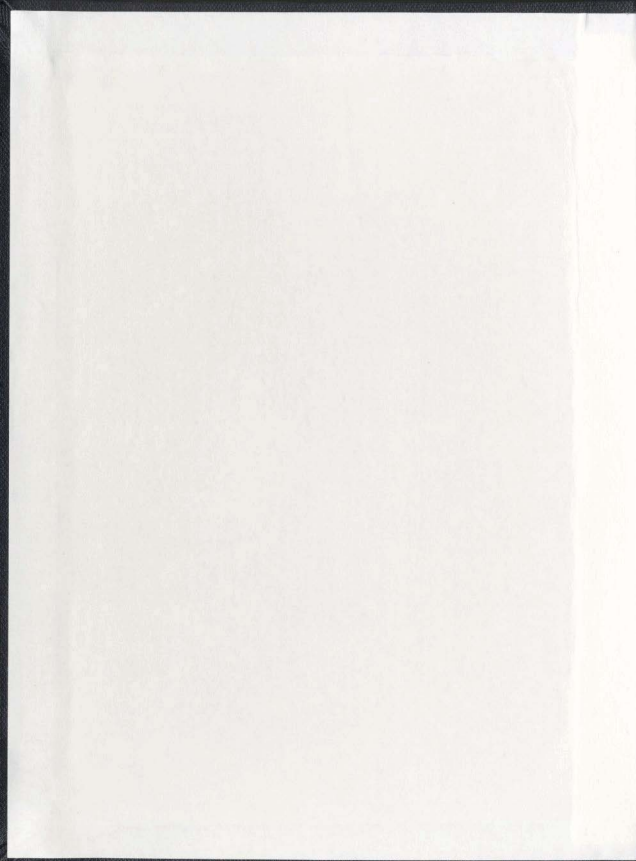
THE CHARACTERIZATION OF REPETITIVE DNA
ELEMENTS IN THE GENOME OF ATLANTIC SALMON
(*Salmo salar*)

CENTRE FOR NEWFOUNDLAND STUDIES

**TOTAL OF 10 PAGES ONLY
MAY BE XEROXED**

(Without Author's Permission)

JOHN L. GOODIER



THE CHARACTERIZATION OF REPETITIVE DNA ELEMENTS
IN THE GENOME OF ATLANTIC SALMON (*Salmo salar*)

John L. Goodier

A thesis submitted to the School of Graduate Studies in
partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Faculty of Medicine
Memorial University of Newfoundland

September 1994

St. John's

Newfoundland



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN 0-612-01857-1

Canada

ABSTRACT

Unrelated families of repetitive DNA were identified in the genome of Atlantic salmon (*Salmo salar*) and characterized.

1) A segment of DNA containing a single *Bgl*I repeat tandem array was isolated from a genomic library. *Bgl*I repeats comprise 2.3% of the *S. salar* genome and were found in the vicinity of rDNA genes. Southern blot hybridization detects a homologue of the Atlantic salmon *Bgl*I repeat in the brown trout (*Salmo trutta*) genome but not in other salmonids. However, a DNA fragment with sequence homology to part of the *Bgl*I repeat has been isolated from Arctic charr (Hartley and Davidson 1994). The *Bgl*I repeat also detects RFLPs in Atlantic salmon.

2) Tc1 transposon-like sequences are present in salmon, trout, and charr species and may belong to several families that do not follow phylogenetic lines. As many as 15,000 copies reside in the Atlantic salmon haploid genome. The complete DNA sequence of one transposon-like element (SALT1) in Atlantic salmon is 1535 base pairs long. It contains terminal inverted repeats and a degenerate open reading frame of 1273 nucleotides whose inferred amino acid sequence shares homology with transposases from *Caenorhabditis* sp. and *Drosophila* sp. Zoo blot analysis indicates that Tc1 transposon-like sequences are present in other lower vertebrates including several fish species and amphibians but the copy number can vary significantly in different lineages.

3) Members of the *Nhe*I family share a great deal of sequence similarity with one another and with corresponding elements present in brown trout (over 95% identity). These elements, which comprise more than 1.2% of the Atlantic salmon

genome, do not exist in long tandem arrays as is typical of satellite DNA. Subfamily structure exists and there is evidence that members of the same subfamilies are found in both *Salmo* species. Sequences homologous to the *NheI* repeat are located within a 1424-bp segment inserted immediately downstream of the 5' end of a *TcI* transposon-like sequence isolated from Atlantic salmon (Radice et al. 1994).

4) Various minisatellite and other short tandem repeat loci were characterized. Southern blot analysis of the phylogenetic distribution of a subset of the minisatellites indicates one sequence to be pervasive among vertebrates, others present only in Salmoninae or Salmonidae species, and one selectively amplified in Atlantic salmon. Their value as probes for detecting variable number tandem repeat loci is assessed. Evidence exists for coevolution of microsatellite and minisatellite arrays at some loci.

LIST OF ABBREVIATIONS

bp, base pair(s); kb, kilobase(s) or 1000 bp; IGS, intergenic spacer; ITR, inverted terminal repeat; LTR, long terminal repeat; Ma, million years ago; NOR, nucleolar organizing region; nt, nucleotide(s); ORF, open reading frame; PFGE, pulsed field gel electrophoresis; PCR, polymerase chain reaction; rDNA, DNA coding for ribosomal RNA; RAPD, randomly amplified polymorphic DNA; RFLP, restriction-fragment length polymorphism; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na₃citrate, pH 7.6; SSM, slipped-strand mispairing; VNTR, variable number tandem repeats.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Dr. William S. Davidson. Past and present members of his lab, have been helpful in many ways, especially S. Bartlett, T. Birt, M. Cutler, V. Drover, V. Friesen, C. Hurst, C. McGowan, H. McVeigh, and S. Sullivan. Technical assistance was provided by S. Rahal and S. Shin. Members of my supervisory committee, Dr. P. Allderdice and Dr. B. Younghusband, gave generously of their time. My special appreciation also to Dr. S. Carr, Dr. K. Egger, and Dr. M. Mulligan for invaluable advice.

Tissue and DNA samples were provided by S. Hartley, V. Pepper, J. Taggart, and F. Yamazaki. The Atlantic salmon genomic library was a gift of J. Wright (Marine Gene Probe Lab., Dalhousie University, Halifax). The pDM238 rDNA probe from *Drosophila melanogaster* was a gift of J. Hancock, Cambridge University, U.K.

Funding for this research was from a National Sciences and Engineering Research Council strategic grant to W.S. Davidson. I am also grateful both to the Faculty of Medicine and the School of Graduate Studies of Memorial University for financial support during the period of this research.

Special appreciation also to my mother for her enduring support and encouragement.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
LIST OF ABBREVIATIONS	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES AND FIGURES	xi
CHAPTER 1. INTRODUCTION	1
1.1 Repetitive DNA in eukaryotes	1
1.2 Interspersed elements	2
1.3 Satellite DNA	5
1.3.1 Types	5
1.3.2 Methods of maintenance and proliferation	7
1.3.2.1 Replication slippage	8
1.3.2.2 Unequal crossing-over	9
1.3.2.3 Amplification	11
1.3.2.4 Gene conversion	12
1.3.2.5 Selection	13
1.4 Repetitive DNA in fish	14
1.4.1 Satellite DNA	14
1.4.2 Interspersed single copy DNA	15
1.4.3 Interspersed tandemly repeated DNA	15

1.5 Practical applications for the study of repetitive DNA in fish	17
1.5.1 Gene-mapping	17
1.5.2 Identification of populations	21
1.5.2.1 Salmonids and Atlantic salmon	23
CHAPTER 2. OBJECTIVES AND APPROACH FOR THE STUDY	26
CHAPTER 3. MATERIALS AND METHODS	32
3.1 DNA isolation	32
3.2 Electrophoresis and Southern blot transfer	32
3.3 Cloning and sequencing of repeat fragments	33
3.3.1 Cloning of the <i>Bgl</i> I repeats	33
3.3.2 Cloning of the <i>Nhe</i> I and <i>Pst</i> I repeats	35
3.3.3 Isolation of a complete SALT1 repeat	35
3.3.4 Isolation and sequencing of the mini- and microsatellites loci	36
3.4 PCR conditions	37
3.4.1 PCR of rDNA genes associated with <i>Bgl</i> I repeats	37
3.4.2 Inter-SALT1 PCR	37
3.4.3 Single-locus DNA PCR	39
3.5 Determination of the phylogenetic distribution of repeat elements	39
3.6 Fingerprint analyses	40
3.7 DNA sequence analysis	41

3.8	Determination of repeat copy number	42
3.9	Cytogenetic techniques	44
3.9.1	Chromosome preparations directly from embryos	44
3.9.2	Chromosome preparations from tissue cultures	45
3.9.3	Chromosome preparations from blood	46
3.9.4	Chromosome banding	47
3.10	Pulsed field gel electrophoresis (PFGE)	47
CHAPTER 4. THE <i>BGLI</i> REPEAT - RESULTS AND DISCUSSION		49
4.1	Cloning of the repeat sequence	49
4.2	Sequencing of pASBg122	54
4.3	Genomic organization of the repeat	62
4.4	Species distribution of the <i>Bgl</i> I repeat	71
4.5	The repeat as a polymorphic population marker	74
4.6	Summary	77
CHAPTER 5. THE SALT1 REPEAT, A Tc1 TRANSPOSON-LIKE ELEMENT - RESULTS AND DISCUSSION		83
5.1	Isolation of a repetitive element identified by <i>Pst</i> I	83
5.2	Characterization of a full-length salmon Tc1 transposon-like element	87
5.3	Species distribution of SALT1 elements	99
5.4	The genomic distribution of SALT1 elements	107
5.5	Summary	113

CHAPTER 6. THE <i>NheI</i> REPEAT - RESULTS AND DISCUSSION	115
6.1 Sequences of Atlantic salmon <i>NheI</i> repeats (ASNHES)	115
6.2 Genome organization of the Atlantic salmon <i>NheI</i> repeat	118
6.3 Intraspecific variation of the <i>NheI</i> repeat	121
6.4 Species distribution of salmon <i>NheI</i> -like sequences	126
6.5 Sequence analyses of <i>Salmo</i> <i>NheI</i> repeats	126
6.6 Summary	136
CHAPTER 7. MINISATELLITE FAMILIES - RESULTS	
AND DISCUSSION	140
7.1 Introduction	140
7.2 The isolation of microsatellite arrays	143
7.3 Different minisatellite families	147
7.4 The junctions of minisatellite arrays - association with other repeat types	153
7.5 Characterization of SsPstIL48	155
7.6 Phylogenetic Distribution of the Minisatellite Sequences	159
7.7 Use of SsBgIIIU20 as a probe for the analysis of population structure	169
7.8 Summary	180
CHAPTER 8. CYTOGENETIC MAPPING OF REPEAT SEQUENCES	182
8.1 Introduction	182
8.1.1 Chromosome banding	182

8.1.2 <i>In situ</i> hybridization	184
8.2 Results and discussion	185
CHAPTER 9. CONCLUSION	194
9.1. Implications of repetitive DNA for species evolution	195
9.2. Organization of repetitive DNA families in salmonids	199
9.3. Future directions	201
REFERENCES	207

LIST OF TABLES AND FIGURES

Table 3.1. PCR primer sequences used in the study.	38
Table 7.1. British and Canadian Atlantic salmon genotypes detected with the minisatellite probe SsBglIIU.20.	179
Fig. 2.1. Southern transfer of PFGE analysis of Atlantic salmon DNA, probed with the insert of pRH3, a plasmid containing a full length copy of the <i>Hind</i> III tandem repeat (Moir 1988).	30
Fig. 4.1. Ethidium bromide stained 1.2% agarose gel of Atlantic salmon DNA digested with restriction endonucleases that produce bands over the background smear.	50
Fig. 4.2. A, left) Atlantic salmon DNA digested with <i>Bgl</i> II, <i>Alu</i> I, <i>Bst</i> EII, <i>Hpa</i> II, and <i>Msp</i> I, and electrophoresed in a 0.8% agarose gel. B, right) Southern blot analysis of gel in A), probed with the 923-bp <i>Bgl</i> II band isolated from an agarose gel.	52
Fig. 4.3. Map of subclone pASBgl22.9, showing the location of the <i>Bgl</i> II repeats.	55
Fig. 4.4. Southern blot analysis of <i>Bgl</i> II partial digests of the salmon DNA insert of subclone pASBgl22.9.	57
Fig. 4.5. A) Sequence of <i>Bgl</i> II repeat variants in pASBgl22.9,	

obtained from M13mp18 or M13mp19 subclones. B) 5'	
flanking region of pASBg122.9. C) 3' flanking region	
of pASBg122.9.	59
Fig. 4.6. Hybridization of the insert of pDM238 with <i>Bgl</i> I repeat	
phage clones (ASBg1 series). A) Intact phage DNA from	
clones identified as containing <i>Bgl</i> I repeats, electrophoresed	
in a 0.8% agarose gel. B) Southern transfer of gel in	
A) hybridized with the pDM238 insert. C) Fragments excised	
from the phage of A) and separated in a 0.7% agarose gel.	
D) Southern transfer of gel in C) hybridized with the pDM238	
insert.	64
Fig. 4.7. Amplification of selected recombinant λ EMBL3 phage DNA	
(ASBg111, 16, 18, 20, 22) and genomic DNA by primers specific	
to rDNA cistron sequence conserved among eukaryote species.	66
Fig. 4.8. PFGE analysis of Newfoundland Atlantic salmon DNA. A) Ethidium	
bromide-stained agarose gel. B) Southern transfer of gel	
in A) hybridized with the insert of <i>Bgl</i> I subclone mp18Alu18.	69
Fig. 4.9. Zoo blot analysis of the distribution of <i>Bgl</i> I elements	
in Salmoninae species, Arctic grayling, and northern pike.	72
Fig. 4.10. Sequence similarity of the Atlantic salmon <i>Bgl</i> I repeat	
(nt 216 to nt 399 of Variant-1, Fig. 4.5) with the Arctic	
charr <i>Dral</i> repeat (ChD:7; Hartley and Davidson 1994).	75

Fig. 4.11. Comparison of <i>Bgl</i> I repeat regions isolated from the Atlantic salmon genomic library.	78
Fig. 4.12. Analysis of genomic DNA from a brown trout (lane 1) and Atlantic salmon from North America (lanes 2 to 13), Scotland (lanes 14 to 17) and Ireland (lanes 18 to 21).	80
Fig. 5.1. Sequences of <i>Pst</i> I fragments of SALT1 repeat elements from various Salmonidae species.	84
Fig. 5.2. Sequence of the full length SALT1 element (SSal1) isolated from an Atlantic salmon genomic library, and comparison of its deduced putative transposase coding region with those of homologous elements.	88
Fig. 5.3. Comparison of homologous sequences within the 5' and 3' inverted repeat sequences of SALT1-SSal1 and analogous 5' sequences from the salmon ependymin gene element.	91
Fig. 5.4. Comparison of homologous sequences of the inverted 5' and 3' repeat sequences flanking fish Tc1-like elements.	94
Fig. 5.5. Alignment of the 5' and 3' flanking sequences of A) ZeffTrana, Tc1-transposon-like sequence from <i>Danio</i> <i>rerio</i> , B) SALT1-SSal1, and C) SmoTc1lt (Radice et al. (1994); #L12206).	97
Fig. 5.6. Zoo blot analysis of the distribution of SALT1-like elements in fish and anuran species.	100

Fig. 5.7. Relationships of the SALT1-like sequences isolated from different Salmonidae species analyzed with the phylogenetic package PHYLIP. The tree in A) was generated with a nearest neighbour-joining algorithm (program NEIGHBOR), and in B) with Fitch-Margoliash and least squares methods (program FITCH).	102
Fig. 5.8. Sequence similarity of SALT1-SSall (nt 1201 to nt 1262, Fig. 5.2) with the Arctic charr <i>Mbo</i> I repeat (ChMb12; Hartley and Davidson 1994; GenBank Acc. #L01073).	105
Fig. 5.9. Autoradiograph of inter-SALT1 PCR of genomic DNA from Atlantic salmon (lanes 1 to 10), other salmonids (lanes 11 to 15), and non-salmonid species (lanes 16 to 23). ³² P-labelled primer a was used alone in the PCR reactions.	108
Fig. 5.10. Autoradiograph of inter-SALT1 PCR of genomic DNA from Atlantic salmon (lanes 1 to 17), other salmonids (lanes 18 to 22), and non-salmonid species (lanes 23 to 30). ³² P-labelled primer c was used alone in the PCR reactions.	111
Fig. 6.1. Comparison of the nucleotide sequences of 16 cloned <i>Nhe</i> I Atlantic salmon repeats (ASNHE) and their corresponding consensus sequence (AS Cons).	116
Fig. 6.2. Genomic organization of the <i>Nhe</i> I repeat. Atlantic salmon DNA from a single individual was digested with	

the restriction enzymes shown and subjected to Southern blot analysis using ASNHE6 as the probe.	119
Fig. 6.3. Southern blot analysis of partial digests of salmon genomic DNA probed with the DNA insert from ASNHE6.	122
Fig. 6.4. Southern blot analysis of salmon DNA from Newfoundland (NF), Nova Scotia (NS), Scotland (SCOT), and Ireland (IRE) digested with <i>HindIII</i> , <i>Sau3A</i> , <i>RsaI</i> and <i>HinfI</i> and probed with the insert of brown trout clone BTNHE68.	124
Fig. 6.5. Zoo blot analysis of the taxonomic distribution of <i>NheI</i> elements. Genomic DNA was digested with <i>NheI</i> and the insert from clone ASNHE6 was used as the probe.	127
Fig. 6.6. Comparison of the nucleotide sequences of 7 cloned <i>NheI</i> Brown trout salmon repeats (BTNHE), their corresponding consensus sequence (BT Cons) shown above, and the ASNHE consensus (AS Cons).	129
Fig. 6.7. Relationships of the ASNHE and BTNHE sequences analyzed with the phylogenetic package PHYLIP. The tree in A) was generated with a nearest neighbour-joining algorithm (program NEIGHBOR), and in B) with Fitch-Margoliash and least squares methods (program FITCH).	131
Fig. 6.8. Comparison of the sequences of part of the Tc1 transposon-like sequence identified by Radice et al.	

(1994)(SmoTc1lt, GenBank Acc. #L12206), the Tc1-like element upstream of the Atlantic salmon cpendymin gene (Ssep; Müller-Schmid et al. 1992), and two copies of the ASNHE consensus sequence.	134
Fig. 6.9. Southern blot analysis of Atlantic salmon DNA digested with the same enzymes as if Fig. 3.1 and probed with the insert of BTNHE68.	137
Fig. 7.1. Sequences of minisatellite arrays isolated from Atlantic salmon.	141
Fig. 7.2. Sequences of repetitive elements isolated from Atlantic salmon.	144
Fig. 7.3. Two similar repetitive elements isolated from Atlantic salmon and containing degenerate (ACGT) _n tetranucleotide and pyrimidine dinucleotide repeats (shown in bold).	148
Fig. 7.4. Sequence of clone SsPstIL.48, isolated from Atlantic salmon. Other similar sequences are shown.	150
Fig. 7.5. PCR amplified alleles of the SsPstIL.48 locus in salmon of Newfoundland, Ireland, and Scotland.	157
Fig. 7.6. Zoo blot analysis for SsBgIIIU.20	160
Fig. 7.7. Zoo blot analysis of the distribution of SsBgIIIL.6 in fish and other vertebrate species.	163
Fig. 7.8. Zoo blot analysis for SsSaclU.44 oligonucleotide probe.	165

Fig. 7.9. Southern blot analysis of Atlantic salmon DNA digested with the same enzymes as in Fig. 4.1 and probed with the oligonucleotide probe of SsSacIU44.	167
Fig. 7.10. Zoo blot analysis for SsAluI16.70	170
Fig. 7.11. Zoo blot analysis for SsPstIL.26.	172
Fig. 7.11. Analysis of genomic DNA from Atlantic salmon of Eastern Canada and Ireland. A) DNA was digested with <i>AluI</i> and probed with SsBgIIIU.20. B) DNA of Atlantic salmon digested with <i>PstI</i> and probed with SsPstIL.26.	174
Fig. 7.13. Analysis of genomic DNA from F1 offspring (lanes 2 to 13) and parents of a female Atlantic salmon X male brown trout hybrid cross. DNA was digested with <i>AluI</i> and probed with the insert of SsBgIIIU.20.	177
Fig. 8.1. Stained metaphase chromosomes from a single primary embryonic tissue culture harvested at 26 days post- fertilization.	187
Fig. 8.2. Stained metaphase chromosomes from A) a single whole embryo, 15 days after fertilization. B) from a single primary tissue culture of a fry 4 months post-fertilization, harvested at 24 days.	189
Fig. 8.3. Stained metaphase chromosomes from a single whole embryo, 14 days after fertilization.	191

CHAPTER 1. INTRODUCTION

1.1 Repetitive DNA in eukaryotes

A large portion of the eukaryote genome is comprised of repetitive, non-coding DNA sequences, ranging in complexity from two to thousands of nucleotides (nt) and existing as both tandem arrays and dispersed copies (Singer 1982). Systems for classifying repetitive DNA are based in part on repetition frequency. Often it is difficult to establish true copy number for a repeat family and it is common to categorize them as of high (in the case of the human genome, 10^5 copies), moderate (10^2 to 10^5 copies) and low (less than 100 copies) repetitiveness. Repetitive components were traditionally studied by reassociation kinetics (C_0t curve analyses)(Britten and Kohne 1968) and isopycnic centrifugation (Waring and Britten 1966). These provide only rough estimates of copy number. Many families have highly divergent members and sequences with less than 70% sequence identity would not be detected under most hybridization conditions. The members permitted in a family therefore will vary according to the hybridization stringency used in the DNA reassociation analyses.

Families of repeat elements are also identified by their sequence complexity and genome dispersion. When the identification is based upon restriction analyses and direct sequence data, a more realistic picture of genome organization is possible than from C_0t curve analyses. Commonly recognized categories include the following:

- 1) multi-copy coding genes such as rDNA and histone gene clusters.
- 2) short arrays of simple motifs, often found in the flanking regions (such as in

promoters and enhancers) or introns of coding genes.

3) dispersed repetitive elements, found as isolated copies flanked by non-related DNA. Included are DNA transposons, retrotransposons (including, in the terminology of Singer (1982), LINES, long interspersed nucleotide sequences), retroposons (including SINES, short interspersed nucleotide sequences) and processed pseudogenes. It is believed that transposition events determine the proliferation of these sequences. These may involve a) direct excision of an element followed by reinsertion into a new location, or b) reverse transcription of an RNA copy of the element.

4) highly repeated sequences existing in tandem arrays. In the case of humans, this group represents at least 10% of the genome (Willard 1989). It includes a) "classical" satellite DNA, and b) short sequence repeats, existing as arrays interspersed throughout the genome, the minisatellites and microsatellites.

1.2 Interspersed elements

The discovery of mobile genetic elements radically altered the view that genomes were static entities passing essentially unchanged from one generation to the next. Transposable elements have been classified depending upon whether their movement within the genome requires an RNA intermediate and reverse transcriptase (class I, retrotransposons or retroposons) or a DNA intermediate (class II, transposons) (Finnegan 1985, 1989).

Unlike retroviruses, retrotransposable elements do not appear capable of transmission between cells, probably because they lack the *env* gene able to encode

envelope glycoproteins. (An *env*-like sequence may exist in the 17.6 element of *Drosophila*; King 1992). Although structurally quite diverse, three major classes have been recognized: 1) those containing long terminal repeats (LTRs) flanked by short inverted repeats of constant size characteristic of each element and resulting from target site sequence duplication. Two main groups have been identified and named after their original *Drosophila* members: the *Copia*-like elements, include Ty of yeast and are also present in flowering plants, and *Gypsy*-like elements, harboured by slime mold, yeast, plants, insects and vertebrates (Xiong and Eickbush 1990; King 1992); 2) those lacking LTRs, and having target site duplications of varying size and an A-rich 3' region (e.g. F, G, Doc, Jockey and I elements of *Drosophila*,ingi of *Trypanosoma*, Tx1 of *Xenopus*, the Cin4 element of maize, and the LINE-1 family of mammals). For reviews see Weiner et al. (1986), chapters in Berg and Howe (1989), and Singer and Berg (1991); 3) some fungal group II mitochondrial introns and a green algae plastid (Lambowitz 1989, Perlman et al. 1989, for review). Doolittle et al. (1989) and Xiong and Eickbush (1990) have explored the phylogenetic relationships and evolution of retrotransposons based on sequence analysis of their reverse transcriptase genes.

The mechanism of transposition for retrogenes or retroposons is unclear, as they themselves do not encode reverse transcriptase and, for mobilization, may require it supplied *in trans* by retroviruses or retrotransposons. This group includes processed pseudogenes which may be copies of mRNAs or, as with the SINE-like elements, small nuclear RNAs such as tRNA or 7SL RNA (Deininger 1989).

Class II mobile elements, or transposons, support their own transposition, but do

not encode reverse transcriptase and probably move through a DNA intermediate. These are typically flanked by short inverted repeats and contain one or more open reading frames (ORFs); included are the insertion (IS) elements of bacteria, P, *hobo*, and *mariner* of *Drosophila* species, Activator-Dissociation (*Ac-Ds*), *Spm* and *Mu* elements of maize, Tam1, Tam2 and Tam3 elements in *Antirrhinum majus*, and Tc1, Tc2 and Tc3 elements in *C. elegans*. Some class II members, generally known as "foldback" elements, are almost entirely inverted repeats: FB elements of *Drosophila*, TU elements of sea urchins, and Tc4 elements of nematodes (see chapters in Berg and Howe 1989; Finnegan 1985, 1989). Although common in bacteria, plants and invertebrates, transposons until recently were unknown in vertebrates. Elements related to the Tc1 family of transposons in nematodes and *Drosophila* have now been found in the genomes of a primitive Agnathan fish (Heierhorst et al. 1992) and modern teleost fish (Henikoff 1992; Radice et al. 1994). It is this group of interspersed repeats which are the concern of Chapter 5.

Although transposable elements have been found in representatives of every order of organism, the distribution of a particular mobile element does not always follow the expected phylogenetic pattern (Robertson 1993). This has complicated classification schemes, which tended to define many small families within class I and class II transposable elements based on the structure of their terminal repeats, their hosts, and their mechanisms of transposition.

Transposons, for example, must encode an enzyme that is required for the integration of the element into the genome. This has provided another means of

classification that is based on a comparison of the genes for the transposases and integrases. For example, sequence similarities among the transposases of the Tc1 elements of *Caenorhabditis*, the *mariner* transposons of *Drosophila*, and IS630 of *Shigella*, indicated that the members of these families of transposons were part of a superfamily (Henikoff 1992). Similarly, the identification of conserved essential amino acids in transposases and integrases allowed families of retroviruses, retrotransposons and IS elements to be linked (Kulkosky et al. 1992). The "D,D35E" motif appears to exist in every integrase and transposase examined to date (Doak et al. 1994), but whether this is evidence for a common ancestor of all transposable elements or convergent evolution of enzyme function is still open for debate.

The observation that a transposable element found in one organism may be absent from closely related organisms but present in organisms that are evolutionarily distantly related is an enigma. This is the case for some transposable elements in insects and these observations have implicated both vertical transmission and horizontal transfer in the evolution of transposons (Daniels et al. 1990; Houck et al. 1991; Flavell and Smith 1992; Robertson 1993; Clark et al. 1994).

1.3 Satellite DNA

1.3.1 Types

Tandemly repeated elements of eukaryote genomes fall into two classes. The first, satellite or "maxi"-satellite was named for its original mode of isolation as bands of different bouyant density segregating from the main body of genomic DNA during

isopycnic centrifugation (Kit 1961). Cross-contamination between bands can occur, however, and the identity of repetitive DNA isolated through centrifugation is best confirmed by sequencing analyses. Six human satellite groups have been obtained and partially characterized (satellite I-IV, α and β ; see review in Vogt 1990).

Satellite DNA includes long arrays of sequence arising from the expansion of monomeric units of considerable length, some over 2 kb (Miklos and Gill 1982). These tend to be concentrated in heterochromatin regions, such as at centromeres and telomeres. Human α -satellite DNA, for example, has arisen from the expansion of a 171-bp basic repeat and is present at the centromere of all chromosomes (Manuelidis 1978; Willard and Wayne 1987). It should be realized that not all heterochromatin is composed of tandemly repeated satellite DNA, nor is all satellite DNA found only in heterochromatin. Satellite DNA is also defined in many sources as being typically late-replicating and non-coding. However, stable RNA transcripts from satellite DNA of species as diverse as newts and humans have been isolated (Epstein et al. 1986; Neuer-Nitsche et al. 1988).

Satellite DNA is usually classified into hierarchies of families and their subfamilies. The complexity can become bewildering; the decision to call a new repeat variant a distinct family by virtue of small sequence divergences, rather than a subfamily of some other group, is often arbitrary. For example, eight centromeric bovine satellites are recognized, although sequence analysis has shown them to be interrelated and possibly evolved from one or two progenitor sequences (Płucienniczak et al. 1982). Some repeat classes are more pragmatic than realistic.

The second class of tandemly repeated elements comprises much shorter arrays of simpler units interspersed throughout the genome. For convenience, this class has been divided into minisatellite DNA (<10-100 bp) and simple sequence repeats or microsatellite DNA (1-6 bp)(Tautz 1993), but the division between the two is arbitrary. Many different motifs have been identified at many chromosomal positions. Simple sequence repeats seem to be more evenly distributed along chromosomes, whereas in humans minisatellites are preferentially located near telomeres (Royle et al. 1988).

Simple sequence or microsatellite DNA, by definition, is distinguished from other repetitive DNA by its shorter and more homogenous sequence motifs, and lack of higher order structures common in satellite DNAs. Some simple sequence DNA (GATA and GACA microsatellites, $(CA)_n$, $(GT)_n$ and $(GA)_n$, $(CT)_n$ arrays, etc.) are found in a wide variety of genomes. It is believed that many arose independently, possibly through errors in replication (see below; Hamada et al. 1982; Levinson et al. 1985).

Short sequence and satellite repetitive DNA elements form a large percentage of the genome of some organisms (over 40 percent in *Drosophila virilis* and approximately 70 percent in the kangaroo rat; Maio 1976). As discussed below, diverse forces may have been involved in their amplification within these genomes.

1.3.2 Methods of maintenance and proliferation

Neither the amount nor the sequence of satellite DNA is well conserved in evolution. In fact, within a single species, chromosome specific subfamilies have arisen through sequence mutation. Superimposed on the basic monomer units of some satellite

DNA families is the "higher-order repeat unit", a large-scale hierarchical organization evolving from the monomers and amplifying through long tandem arrays (a concept proposed originally by Southern (1975), and reviewed for human α -satellite by Willard and Wayne (1987)). Different subfamily arrays can often be detected by signature loss or gain of tandem restriction sites from specific blocks of the repeat family. The chromosome specificity of at least some of these arrays, studied extensively in humans (Willard 1989), is presumably a feature of other genomes.

Both human X and Y chromosomes contain α -satellite subfamilies with "higher-order repeat units" of approximately 2 kb in length based on repeated multimers of a 171-bp prototype sequence (Laursen et al. 1992). Possibly each human chromosome has a distinct α -satellite super-repeat, as defined by restriction-site organization and DNA sequences (Jorgenson et al. 1986; Willard and Wayne 1987; Choo et al. 1991). The challenge to biologists is to explain how such "unity in diversity" has arisen. A number of explanations have been offered, but no satisfactory unifying hypothesis has yet been achieved.

1.3.2.1 Replication slippage

The copy number of minisatellite and microsatellite DNA can vary greatly both between species and between individuals of a single species. (Such variation is the cause of the many restriction fragment length polymorphisms (RFLPs) detected first in humans and then in many other organisms, and forms the basis of the techniques of DNA profiling or DNA fingerprinting (see below; Jeffreys et al. 1985b). Some

researchers (Tautz et al. 1986; Levinson and Gutman 1987) propose that the major mechanism involved in the evolution of microsatellite DNA is slipped-strand mispairing (SSM) at replication. There is direct evidence that slippage operates on microsatellites (Schlötterer and Tautz 1992).

Short runs of simple sequence will often occur by chance in a genome. Initial amplification by replication strand slippage will expand the substrate and hence the chance that SSM will occur in the future. Single nucleotide substitutions, deletions or insertions may convert one simple sequence to another. Human satellite II, for example, is a degenerate form of the pentameric repeat 5'-TTCCA-3', which is also the basic unit of Satellite III (Prosser et al. 1986). The latter is the older in evolutionary time, being present in man and higher primates, while the former is unique to humans (Mitchell et al. 1981).

1.3.2.2 Unequal crossing-over

Once localized duplications of short stretches of DNA have occurred, they may be amplified or dispersed through the genome by recombination events involving non-allelic homologous repeats. This unequal crossing-over may occur between repeats on sister chromatids or homologous chromosomes. Occurring at meiosis or pre-meiotic mitosis, unequal recombination increases or decreases copy number in progeny. New array lengths may eventually become fixed in a population by genetic drift. In fact, computer models have shown that tandem repeats undergoing unequal recombination will tend to evolve toward higher order periodicities (for example, Smith 1976; Stephan

1989). At high rates of recombination short repeats similar to minisatellites arise with little higher order structure. When recombination rates are maintained at a low level, however, long and more heterogeneous repeat patterns are generated.

It is considered that chromosome-specific α -satellite subfamilies arose through homogenization induced by unequal exchange between homologous chromosomes (Willard and Waye 1987). On the other hand, satellite DNA of the centromere and short arms of the acrocentric human chromosomes are not chromosome specific, and show considerable interchromosomal homogeneity (Waye and Willard 1989). This observation suggests that homologous recombination may be restricted to certain regions of the centromere, or that non-homologous chromosomal recombination may occur (Choo et al. 1991; Vissel and Choo 1992).

It is probable that some families of satellite DNA derive from a single "founder" sequence. Invoking hypotheses of SSM and unequal crossing-over alone, it is difficult to explain the proliferation at distant chromosomal locations of related subfamilies. Furthermore, although unequal crossing-over alone will not alter the overall number of repeat copies within a population (each cross-over event results in a gain of repeats on one strand and accompanying loss on another), intrastrand exchange will always result in a decrease in copy number. Mathematical models presented by Walsh (1987) show that, in the absence of selection or amplification, arrays will be lost over time. He reviews experimental evidence for actual loss by intrastrand exchange, including the attrition of alphoid satellite DNA from human fibroblasts, chromosome diminution as observed in certain species, and loss of satellite sequences from plasmids replicating in

Escherichia coli. Furthermore, the existence of extrachromosomal, circular fragments of alphoid DNA, strongly suggests *in vivo* occurrence of intrastrand recombination (Jones and Potter 1985; Okamura et al. 1987; Rossi et al. 1990).

1.3.2.3 Amplification

A scenario of extensive and continued unequal recombination events building large and complex repeat families is at variance with the genotypic stability, the low rate of recombination of heterochromatin (the domain of much satellite DNA), and the low mutation and translocation rates characteristic of mammalian genes and genomes. Some researchers, therefore, have suggested that saltatory amplification events have played a more important role in the evolution of some satellite and simple sequence DNA families. Amplification is defined here as the increase in copy number of a DNA sequence in a single replication event. Gene amplification is a normal process in some organisms and suggests mechanisms by which tandem arrays may be created and expanded. In *Xenopus*, for example, extrachromosomal rDNA copies arise in the germ cell and are amplified many-fold by rolling circle replication (Dawid et al. 1970). Circular plasmids of repetitive DNA (noted above), if they possess an origin of replication, could be candidates for rolling circle replication. Homologous recombination could then reintegrate the amplified repeats into a chromosome. Another mechanism of amplification might begin with multiple "firing" of a nearby replication origin, amplifying an array of repeats in a manner analogous to the "onion-skin model" of *Drosophila* chorion gene amplification (Wakimoto et al. 1986). Subsequent

recombination could weld the different strands into a single long tandem array.

1.3.2.4 Gene conversion

Within some repeat families, both tandemly repeated and interspersed, the variation expected to occur between members as a result of non-selective, chance mutations and that actually observed are quite different. For example, LINE-1 sequences are found in most mammalian species and retain considerable homology to each other over much of their length (Singer and Skowronski 1985).

The fixation of repeat units within a species is part of a process called "molecular drive" (Dover 1982). It seeks to define mechanisms which preserve relative sequence homogeneity of repeated sequence families. One such mechanism is gene conversion. Gene conversion is defined as a non-reciprocal transfer of genetic information: sequence from one region of the genome is used as a template to replace a similar sequence at another location. Molecular studies have shown the phenomenon to occur within a single chromosome and between homologous and nonhomologous chromosomes and to involve up to several kilobases of DNA (reviewed in Dover 1982). While clearly demonstrated for interspersed elements (Kass et al. 1994, and references therein), the occurrence of gene conversion within tandem and high-copy repeat families is still speculative. However, if the process in a repeat family is biased toward a small number of donor sequences, or "molecular drivers", converting a large number of recipients, homogenization of the family could occur.

1.3.2.5 Selection

Although it is possible to explain the expansion and contraction of arrays of repeats in terms of the forces of molecular drive and DNA metabolism alone (e.g., replication slippage, unequal crossing-over, amplification and gene conversion), it is reasonable that selection acts to maintain at least a minimum number of copies of arrays which produce necessary gene products, such as rDNA or histone genes. Selection may be less strict on imposing an upper limit on copy number, although undoubtedly overproduction of some gene proteins would be destabilizing.

While copy numbers of individual repeat families may vary, even among individuals of a single species, the overall amount of repetitive DNA remains fairly constant within a species. The fact that repetitive DNA tends neither to be lost from nor overpopulate a genome suggests that some, as yet undefined, function encourages its selective maintenance. Roles for satellite sequence DNA in gene expression, modulation of mRNA stability, nucleosome spacing, recombination, chromatin condensation, and homologous chromosome recognition have all been proposed (see Vogt 1990, and references therein). Proteins have been found that specifically bind satellite sequence, including simple sequence DNA (Masumoto et al. 1989; Radic et al. 1992, Yee et al. 1991). Interspersed mobile elements, considered by some as molecular parasites (Orgel and Crick 1980; Doolittle and Sapienza 1980) may have implications for species evolution. This concept is elaborated in the conclusion (Section 9.1).

Obviously copy number is not strictly controlled, although an upper limit may be indirectly imposed by the sheer stress which a vastly expanded genome would place

on the replication machinery of the cell. Furthermore, Stephan and Cho (1994) found that when natural selection, as a constraint parameter limiting sequence length but not base pair order, was introduced to computer simulation models of tandem repeat growth, the system tended to be driven toward the formation of periodic patterns. This tendency to self-organization into tandem repeat blocks favoured shorter sequences and encouraged their maintenance in the genome: by recombining more often, sequences having a periodic organization were able to escape elimination.

1.4 Repetitive DNA in fish

1.4.1 Satellite DNA

Satellite-like tandemly repeated DNA has been isolated from carp (*Cyprinus carpio*; He et al. 1987; Datta et al. 1988; Huang et al. 1993, this latter actually being misidentified 5S RNA, W.S. Davidson and V. Drovot, personal communication), pollock (*Pollichus virens*; Denovan and Wright 1990), Japanese hagfish (*Eptatreis okinosanensis*; Kubota et al. 1993), sculpin (*Batrachocottus baikalensis*; S.Y. Slobodanyuk, unpub., GenBank Acc. #X67745), a cyprinid (*Nothops lutrensis*; McFar et al. 1988), sheephead minnow (*Cyprinodon variegatus*; Turner et al. 1991; Elder and Turner 1994), sea bream (*Sparus aurata*; Garrido-Ramos et al. 1994); Cichlidae (*Oreochromis* sp., *Protonotus similis*, *Sarotherodon* sp., *Tilapia* sp.; Wright 1989; Franck et al. 1992; Franck 1993; Franck and Wright 1993), zebrafish (*Danio rerio*; Ekker et al. 1992; He et al. 1992), Arctic char (*Salvelinus alpinus*; Hartley and Davidson 1994), rainbow trout (*Oncorhynchus mykiss*; Moir 1988), and sockeye salmon

(*O. nerka*; Borkhsenius and Chernov 1988), and includes an 800-bp repeat characteristically cut by *Hind*III and well conserved in salmonids (Moir 1988).

1.4.2 Interspersed single copy DNA

Interspersed elements have been isolated predominantly from the genomes of salmonid fish. These harbour several types of retroposon-like sequences, including retroviral LTR homologues (Moir and Dixon 1988), a long repeat from rainbow trout identified by *Bam*HI and resembling mammalian LINE sequences (Moir 1988; Winkfein et al. 1988), and tRNA-derived SINE-like elements from Pacific salmonids (Kido et al. 1991). Several studies have reported retroviral particles in fish (summarized in Hedrick and Wingfield 1991, and Stuart et al. 1992). The first characterization of retrotransposon gene sequences was by Stuart et al. (1992) of a *pol* gene fragment from trout, and Flavell and Smith (1992) who succeeded in amplifying by polymerase chain reaction (PCR) a portion of a reverse transcriptase gene characteristic of the Ty1-Copia retrotransposon group. The recent discovery in fish of elements similar to Tc1 transposon-like sequences hitherto found only in nematodes and *Drosophila* will be dealt with in Chapter 5.

1.4.3 Interspersed tandemly repeated DNA

Much attention has focussed on the characterization of microsatellite loci and the design of flanking primers for PCR analysis. Sequence information has been obtained for microsatellites from zebrafish (Goff et al. 1992), Atlantic salmon (Slettan et al.

1993), brown trout (Estoup et al. 1993), and Atlantic cod (*Gadus morhua*) and rainbow trout (Brooker et al. 1994). Minisatellites have received less attention. DNA fingerprinting studies have depended largely on mammalian, *Drosophila*, or M13 bacteriophage DNA probes which detect certain pervasive variable number tandem repeat (VNTR) families (Wright (1993) for review).

Many human minisatellite loci have been isolated and their nucleotide sequences determined (Inglehearn and Cooke 1990; Armour et al. 1992). Additional sequence data have been obtained from primates (Royle et al. 1994), mouse (Kominami et al. 1988; Kelly et al. 1989), cattle (Kashi et al. 1990), pig (Coppietiers et al. 1990; Brenig and Brem 1991; Kitazawa et al. 1994), bird (Gyllensten et al. 1989), insects (Blanchetot 1991; Jacobson et al. 1992; Paulsson et al. 1992; Hankeln et al. 1994), plants (Broun and Tanksley 1993; Winberg et al. 1993), and some fish. The sequence of a single locus has been described for tilapia (*Oreochromis niloticus*; Bentzen et al. 1991). A second from Atlantic salmon has been characterized (Bentzen et al. 1991; Bentzen and Wright 1993), while Taggart and Ferguson (1990a) report the isolation of four minisatellite loci from *S. salar* and Prodöhl et al. (1994) five from brown trout (*S. trutta*), but provide no sequence analysis. All of these were isolated by screening genomic libraries with the human-derived DNA fingerprinting probes, 33.15 or 33.6 (Jeffreys et al. 1985a,b). Brenner et al. (1993), using a random sequencing strategy, identified 16 minisatellite types from the tetraodontoid pufferfish (*Fugu rubripes*), but the sequence of only one of these is available. This repeat is 118 bp, longer than that typically classed as "mini"-satellite DNA.

1.5 Practical applications for the study of repetitive DNA in fish

Despite the many questions which remain concerning the origin and genomic function (if any) of non-coding repetitive DNA, its study is providing population geneticists and evolutionary biologists with some powerful tools. In the case of fish, hypervariable simple sequence satellite DNA sequences have been used for testing of parentage and mating strategies (Rico et al. 1991; Schartl et al. 1993; Gross et al. 1994), identification of species markers (Heath et al. 1993), and monitoring the success rate of gynogenesis induction (Carver et al. 1991). Satellite repeat sequences have been found to be sex-linked in certain fish, and so may prove useful in aquaculture (Borkhsenius and Chernov 1988; Lloyd et al. 1989; Nanda et al 1990; Devlin et al. 1991). Sequence data of more conserved repetitive elements are being applied to phylogenetic reconstruction in salmonids (Kido et al. 1991; Murata et al. 1993; Phillips and Pleyte, 1991; Phillips et al. 1992; Franck et al. 1994). Recently the possibility that interspersed repetitive sequences may facilitate integration and mapping of transgenes has been studied (He et al. 1992).

The application of repetitive DNA to problems of gene mapping and the analysis of population structuring holds special promise. It was to develop DNA probes for ultimate use with salmonids in these areas that the present project was initiated.

1.5.1 Gene mapping

The human genome project will have several spin-offs, among them being the development of new techniques and approaches that will allow other vertebrate genomes

to be examined. Comparative biochemical studies have always been recognized as the means for determining what structures and mechanisms are shared by groups of organisms and what is unique to a particular species. Such studies allow unifying theories in biology to be developed. So will it be with comparative genome mapping.

Fish are the most primitive of the vertebrate classes. As such, an investigation of fish genome organization and construction of fish genetic maps will provide a more realistic data base with which the genomes of higher vertebrates can be compared. In addition, genetic mapping of commercially important aquaculture species should increase the efficiency of breeding programs through the use of marker-assisted selection (Lande and Thompson 1990). This procedure involves comparing information derived from genetic polymorphisms (marker loci) with results of phenotypic variation and then detecting associations between certain marker loci and quantitative trait loci (e.g. disease resistance, increased growth rate, late maturation). Considerable progress has already been made in plants by this means (Paterson et al. 1991). Moreover, the detection of polymorphic loci, a step that is essential for gene mapping, will provide the tools for discriminating between different populations of fish and for assessing the impact of domesticated escapees on wild populations (see below).

Significant progress in gene mapping has only occurred in fish from four orders: Cyprinodontiformes (mummichog of the family Cyprinodontidae (genus *Fundulus*), medaka of the family Oryziidae (genus *Oryzias*), and livebearers of the family Poeciliidae which include common aquarium fish such as guppies and mollies (genus *Poecilia*), and platyfish and swordtails (genus *Xiphophorus*), and the confamilial genus

Poeciliopsis); Cypriniformes (zebrafish, family Cyprinidae); Salmoniformes (salmon, trout and charr); and Perciformes (represented by sunfishes, family Centrarchidae) (for review see Goodier and Davidson 1993a).

May and Johnson (1990) have constructed a composite "salmonid gene map", involving more than fifty loci, by combining all the linkage analyses derived from breeding studies involving intra- and inter-specific crosses of several species in three genera (*Salmo*, *Oncorhynchus*, and *Salvelinus*). Most of the loci that have been mapped in fish were positioned by traditional genetic linkage analysis and encode enzymes detectable by histological activity stains (with the exception of the tumor-inducing and pigment loci of Poeciliid fish). Polymorphisms in these enzymes are usually identified by starch gel electrophoresis and much of the gene mapping in fish species is an extension of studies on their population structure. Continued studies based solely on protein coding loci will not greatly increase the saturation of the poeciliid and salmonid gene maps. Many previously untested enzymes are difficult to resolve or detect by techniques of protein electrophoresis. Silent alleles, duplicate loci, and tissue specific expression confound the mapping of some genes. In some fish groups, such as salmonids, the heterozygosity of protein loci is low (Gyllenstein 1985; Ryman 1983).

Nevertheless, current maps of poeciliid and salmonid genomes now provide sufficient anchor loci for rapidly assigning new sites. The mapping of genes in mammalian systems increased in rapidity with the use of nuclear RFLPs as markers. The analysis of RFLPs usually involves digestion of genomic DNA with restriction endonucleases, separation of the fragments by agarose gel electrophoresis, followed by

Southern blotting and hybridization with a DNA probe, usually a cloned fragment. Useful DNA probes will detect differences in restriction fragment lengths arising from loss or gain of enzyme recognition sites or from deletions or insertions of stretches of DNA between sites (Botstein 1980).

The use of multiple allelic markers, involving repetitive genomic DNA sequences are proving more useful than diallelic RFLPs in mammalian studies. These include variable number of tandem repeats (VNTR) sequences as well as interspersed repeats, such as mammalian LINE and SINE elements. These have been applied with much success to mapping new loci in the mouse genome using inter-specific backcrosses (Cox and Lehrach 1991). Some approaches have employed RFLP analysis and Southern blotting with low copy number gene probes that identify multiple allelic loci, the segregation of which could be followed through familial generations (Nakamura et al. 1987; Siracusa 1991).

RFLP analysis, however, is not without its drawbacks. It requires significant amounts of DNA, usually necessitating small organisms be sacrificed. Analysis by Southern blotting is time-consuming and the identification and isolation of informative probes is often tedious. Gene mappers are, therefore, showing increasing interest in the powerful PCR technology which permits amplification of any sequence of interest from nanogram amounts of a DNA sample. One area of application seem especially promising: PCR amplification of genomic "hypervariable" repetitive elements, notably simple sequence mini- and microsatellite loci. Two approaches involve use of 1) primers designed to amplify specific VNTR regions from unique flanking sequences

(Love et al. 1990), or 2) primers directed outward from interspersed repeats which amplify intervening single copy sequence. In either case, PCR products may be directly visualized as bands on ethidium bromide-stained gels. Interspersed repetitive sequence PCR (IRS-PCR) has used primers homologous to SINE and LINE sequences: Alu repeats in humans (Ledbetter et al. 1990; Sinnet et al. 1990) and B1, B2 and L1 repeats in the mouse (Cox and Lehrach 1991; Cox et al. 1991).

A third approach, applies the discovery that short randomly generated PCR primers can be used to reproducibly amplify segments of DNA from a large number of species (Welsh and McClelland 1990; Williams et al. 1990). Unlike VNTR-PCR mapping, no prior knowledge of genome sequence is required. The randomly amplified polymorphic DNA (RAPD) products may be separated on an agarose or polyacrylamide gel and directly visualized by ethidium bromide staining or by radiolabelling and autoradiography. Polymorphic segments are frequently detected. Using 401 RAPDs and 13 simple sequence repeats, Postlethwait et al. (1994) have generated a 5.8 centiMorgan (cM) linkage map for the zebrafish.

1.5.2 Identification of populations

The use of repetitive DNA sequence holds potential value for fine scale taxonomic or population-level investigations. In the case of maxi-satellite and interspersed repetitive DNA, the united action of unequal crossing-over, gene conversion and, in some instances, replicative transposition, collectively known as "molecular drive", can act to homogenize members of each repeat family in a non-

random, non-Mendelian manner (Section 1.3.2.4). This phenomenon of "concerted evolution" (Arnheim 1983) has been most evident at the species level. Such intra-specific homogeneity would not be expected if repeat elements were evolving independently.

Recent evidence suggests that the processes of "molecular drive" can also act at the population level. Sequence-based studies of pupfish by Elder and Turner (1994) and Turner et al. (1991) involved gel isolation and cloning of anonymous maxi-satellite DNA. The average sequence divergence of a number of monomer units of a satellite DNA family was assessed at the intra-population level and found to differ significantly between known populations. Sequence variation in maxi-satellite DNA has also been applied to population studies of cetaceans (Amos and Dover 1991). The possible existence of subfamilies, however, demands that a large number of individual repeat units be examined in order that a representative sample of the whole repeat family may be obtained (Pascale et al. 1993). If paralogous loci are excluded from the analysis, and orthologous loci only (perhaps amplified by PCR from flanking sequence) are compared between putative populations, fewer sequences may be examined with increased reliability of the results.

Not only the degree of sequence similarity, but also the overall unit copy number of some satellite repeat types (as determined by dot blot hybridization studies) can vary between populations and hence be diagnostic. For example, both the *Sau3A* repetitive element and rRNA gene cistrons differ in copy number between anadromous and dwarf forms of sockeye salmon of the Lake Dalneye, Kamchatka population (Borkhsenius and

Chernov 1988).

The high mutation rates of simple sequence repetitive DNA has been seen to limit the usefulness of VNTR polymorphisms as population markers. Laughlin and Turner (1994) point out that most studies to date have been limited to small or inbred populations with very little gene flow (including invertebrates (Hauser et al. 1992), fish (Turner et al. 1990; Wirgin et al. 1991), and mammals (Varvio and Kaukinen (1993); Buchanan et al. 1994)). However, as Laughlin and Turner (1994) demonstrate for sailfin molly (*Poecilia latipinna*), by judicious choice of VNTR probes showing sufficiently low levels of variation, multilocus DNA fingerprinting may be applied to the analysis of pervasive outbred populations.

1.5.2.1 Salmonids and Atlantic Salmon

Approximately 70 species of salmonids are contained within three different subfamilies: Coregoninae, Thymallinae, and Salmoninae. Salmoninae are divided into seven genera: *Brachymystax*, *Acantholingua*, *Salmothymus*, *Hucho*, *Oncorhynchus*, *Salmo* (Atlantic salmon and brown trout), and *Salvelinus* (Nelson 1994). Salmonid taxonomy was originally defined by studies of morphology and meristics, and more recently by karyology, DNA-DNA hybridization, protein allozyme variability, and mitochondrial DNA sequence variability (Thomas and Beckenbach 1989; Grewe et al. 1990, and references therein).

The ancestor of present day salmonids (salmon, trout and charr) underwent a genome duplication event between 25 and 100 million years ago (Ohno 1970; Ohno et

al. 1969). The genomes of extant salmonids have become approximately three-quarters diploidized but retain evidence for the double complement of chromosomes in the expression patterns of many of their duplicated genes (Allendorf and Thorgaard 1984). The economic importance of these species coupled with evolutionary interest in how they have adapted their duplicated genome has resulted in many genetic studies on salmonids. Many of these studies involved population genetic analysis using allozyme markers which confirmed the tetraploid history and led to the use of protein polymorphisms in linkage analysis.

More pragmatically, the economic importance of Atlantic salmon in commercial fisheries, as a sport fish and in aquaculture, has stimulated a considerable number of genetic studies. The main drive has been to obtain genetic markers that can distinguish and monitor different populations (Hallerman and Beckmann 1988). The ability to do so would have immediate applications. Prior to a moratorium imposed in 1993, fleets from several nations participated annually in the valuable sea fishery off the coast of West Greenland, with Atlantic salmon from both Europe and North America contributing to the harvests. It is important for both political and conservation reasons, therefore, that the relative contribution of populations from the two continents be considered when allocating optimal harvest quotas.

In aquaculture better genetic markers are needed for two reasons: 1) to monitor the effects that escapees from culture pens might have on native populations, and 2) to monitor loss of genetic variation within hatchery stocks. It is now a basic tenet of conservation ecology that genetic diversity is important to the health of a species. This

demands that different genetic lineages within a species be monitored and maintained.

Until recently the techniques that were used to search for these population markers concentrated on scale ageing, meristics, protein electrophoresis, mitochondrial DNA RFLP analysis, and more recently mitochondrial DNA sequencing (Bermingham et al. 1991; Davidson et al. 1989). Researchers have frequently commented on the low level of heterozygosity in Atlantic salmon detectable by these methods. Three main groups of Atlantic salmon are indicated: one in North America, another in western Europe, and a third in the Baltic (Stahl 1987; Bermingham et al. 1991). However, there is evidence to suggest that even these groups are not clearly defined and the delineation of local population substructuring remains elusive (McVeigh et al. 1991). Power (1981) has estimated that more than 500 populations (or stocks) of anadromous Atlantic salmon exist along the shores of Newfoundland and the Quebec-Labrador peninsula alone. Salmon have recolonized the North Atlantic only since the last glaciation 8,000 to 10,000 years ago, and therefore population-specific markers have had a very short time in which to evolve.

In an attempt to obtain more discriminating markers, several researchers have turned their attention to repetitive elements in the salmon genome. These have included the genes for ribosomal RNA (Cutler et al. 1991), and mini- and microsatellites (Taggart and Ferguson 1990a,b; Bentzen et al. 1991, 1993; Slettan et al. 1993).

CHAPTER 2. OBJECTIVES AND APPROACH FOR THE STUDY

The objectives of this study were to isolate and characterize as many types of repetitive elements as possible from the genome of Atlantic salmon. In the process, inferences concerning genomic organization in this species were to be made. It was also anticipated that the different repeat types would provide a set of tools which might eventually be turned to solving questions concerning salmonid phylogeny, population structuring, and gene mapping. With this in mind, certain preliminary tests of the efficacy of these repeats were conducted, but it was considered that detailed studies were outside the scope of this project.

As determined by reassociation kinetics, it is estimated that 60% (Gharrett et al. 1977; Hanham and Smith 1980) or more (Schmidtke et al. 1979) of the salmonid genome is composed of repetitive DNA, but highly repeated or satellite DNA has not been identified by isopycnic gradient centrifugation (Bernardi and Bernardi 1990). A useful approach for revealing repetitive DNA involves digesting genomic DNA with a restriction endonuclease, separating the fragments in an agarose gel, and staining with ethidium bromide. A bright band against the background smear indicates the presence of a repeat element, either tandemly arrayed and containing a single recognition site, or interspersed and containing two sites.

In the present study, the above approach was used for the initial isolation of repetitive DNA from the genome of Atlantic salmon. Some individual loci were also isolated from an Atlantic salmon DNA genomic library cloned in bacteriophage λ EMBL3 vector. All methods and materials used to isolate and characterize the

different elements are presented in Chapter 3. Each type of repeat element is allotted its own combined Results and Discussion section as organized in the following chapters. These precede a final concluding chapter. Much of the work in Chapters 4 to 7 has been published or is currently under review, as follows:

Chapter 4: A tandemly arrayed element identified by *Bgl*I with basic monomer unit over 400 bp long with some copies found near rDNA genes (Goodier and Davidson 1993b);

Chapter 5: An interspersed element (designated SALT1-SSal1, SALmonid Transposon of *S. Salar*) with an internal sequence fragment identified by *Pst*II, and showing similarity to the Tc1 family of transposons, originally described in nematodes and *Drosophila* (Goodier and Davidson 1994a);

Chapter 6: A repetitive element identified by *Nhe*I, possibly existing in short tandem arrays rather than the long tracts typical of satellite DNA (Goodier and Davidson 1994b);

Chapter 7: Various minisatellite and microsatellite loci isolated either incidentally from the background of visible restricted DNA bands (eg. those generated by *Bgl*II, *Pst*I, and *Sac*I) or from the genomic library (Goodier and Davidson 1994c).

As part of the original aims of this project, it was anticipated that information about the large-scale genomic organization of the different repeat types might be

provided by two different techniques:

1) Chromosomal distribution would be directly visualized by fluorescent in-situ hybridization (FISH) of biotin-labelled repetitive DNA probes to salmon metaphase chromosomes. By this technique, amplification and detection of the probe signal is enhanced during successive applications of avidin-conjugated fluorescein and biotinylated anti-avidin. The chromosomes themselves are counterstained with propidium iodide, and visualized under a fluorescent microscope. Chromosomes appear red, and the site of the hybridized probe yellow.

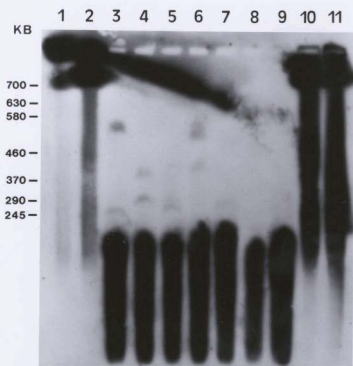
Success was predicated on the ability to consistently obtain fresh metaphase preparations of reasonably high mitotic index. Failure to accomplish this prevented application of the in-situ protocol. An initial feasibility study did test a variety of procedures for preparing mitotic chromosomes and for chromosome banding (Section 3.9). Chapter 8 reviews literature dealing with chromosomal banding and gene localization in fish, followed by a brief discussion of results and problems encountered during the present study.

2) Pulsed field gel electrophoresis (PFGE) permits the separation of very large fragments of DNA in agarose gels placed within an alternating electrical field (see Cantor et al. (1988) for a review of the principles involved). Preliminary studies suggested that the technique might be used to infer spacing between interspersed repeats and different tandem repeat arrays. For example, I examined patterns of hybridization of the probe pRH3, a plasmid containing the *Hind*III repeat of Atlantic salmon (a gift of R. Moir; see section 1.4.1). Salmon DNA was digested with various rare-cutting

enzymes and subjected to PFGE and Southern blotting. The enzymes *MluI* and *XhoI* generate a high molecular weight smear without obvious bands (Fig. 2.1). This smear is not seen for *KpnI* and *NheI*, although in the case of *KpnI*, a few distinct high molecular weight bands show RFLP variation between individuals. These likely represent allelic variants of individual loci.

It became apparent, however, that in addition to elaborate sample preparations and very long electrophoresis times, the time required for optimization of the PFGE protocol might prove considerable. PFGE, therefore, was not extensively used in this project. Procedures are briefly described in Section 3.0.

Fig. 2.1. Southern transfer of PFGE analysis of Atlantic salmon DNA, probed with the insert of pRH3, a plasmid containing a full length copy of the *Hind*III tandem repeat (Moir 1988). Electrophoretic conditions were 40 hr at pulse intervals ramped from 60 sec to 100 sec, with a constant temperature of 12.5°C and current of 105 V. Restriction enzymes were: lane 1, unrestricted DNA; lane 2, unrestricted DNA, incubated in restriction buffer; lanes 3 to 8, *Kpn*I; lane 9, *Nhe*I; lane 10, *Mlu*I; lane 11, *Xho*I. Atlantic salmon were from: lanes, 1 and 2, N.E. Placentia, NF; lanes 3 to 5, Gambo Pond, NF; lane 6, Exploits R., NF; lane 7, N.E. Placentia, NF; lane 8, St. John R., New Brunswick; lanes 9 to 11, N.E. Placentia, NF. Hybridization was at 42°C as described in section 3.3.1, and washing was at high stringency at 55°C. Sizes of markers shown at left are *Saccharomyces cerevisiae* whole chromosomal DNA.



CHAPTER 3. MATERIALS AND METHODS

3.1 DNA isolation

Much of the DNA used in this project was drawn from a large multi-species collection previously extracted in the W.S. Davidson lab and stored in distilled water or TE buffer (10 mM TrisCl, pH 8.0/1mM EDTA, pH 8.0) at 4°C. Atlantic salmon genomic DNA was extracted from whole blood, liver, or muscle as described in Cutler et al. (1991). DNA from other species was obtained by standard procedures (Sambrook et al. 1989; Miller et al. 1988). DNA extractions made specifically for this project were according to Taggart et al. (1992).

3.2 Electrophoresis and Southern blot transfer

Electrophoresis of DNA was carried out in 1× TBE buffer (0.89M Tris-borate/1mM EDTA, pH 8.3). All electrophoresis intended for the purpose of fragment isolation used 1× TAE buffer (0.04M Tris-acetate/1mM EDTA, pH 8.0). After staining with ethidium bromide (Sigma), and in preparation for Southern transfer, gels were treated three times in 1.5 M NaCl/0.5 M NaOH for approximately 30 min each time at room temperature. The transfer solution was 1.5 M NaCl/0.25 M NaOH. Gels were blotted overnight by capillary action to Hybond N (Amersham) or BioDyne (Pall) nylon membranes. Membranes were baked at 80°C for at least 1 hr and stored at room temperature.

3.3 Cloning and sequencing of repeat fragments

3.3.1 Cloning of the *Bgl*I repeats

To isolate the *Bgl*I repeat, genomic DNA from a single Newfoundland Atlantic salmon individual was digested with the restriction endonuclease *Bgl*I according to the manufacturer's instructions (Pharmacia) and the fragments were separated by electrophoresis in a 0.8% agarose gel in TAE buffer. The 923-bp *Bgl*I band was extracted (Geneclean, Bio/Can Scientific, Ontario), labelled with [α - 32 P]dCTP (Amersham) and a Random Primers DNA Labeling System kit (Bethesda Research Lab., Ontario), and used as a crude probe to screen an Atlantic salmon DNA genomic library cloned in λ EMBL3 vector (gift of J. Wright, Marine Gene Probe Lab, Dalhousie University, Halifax). To accomplish this, *E. coli* NM539 cells were infected with library phage, mixed with 8 ml of 0.7% top agarose, and poured onto 137 mm diameter culture plates containing 1.5% agar in LB medium (1.0% bacto-tryptone (Difco)/0.5% bacto-yeast extract (Difco)/1.0% NaCl). Circular Hybond N nylon membranes with plaque replicates were sealed in plastic freezer bags together with 10 ml of 6 \times SSC/50% formamide/5 \times Denhardt's solution (0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone)/50 mM phosphate buffer, pH 6.5/0.5% SDS and prehybridized for 3 to 5 hr at 42°C in a water bath. Each membrane was then hybridized in 5 ml of the same solution and 2 \times 10⁸ dpm of denatured probe for 20 hr at 42°C. (A Sephadex G-50 (Pharmacia) column was used to remove unincorporated nucleotides from all labelled probes used in this project). Washing was at high stringency (twice in 2 \times SSC for 20 min at room temperature; 4 times in 0.1 \times SSC/0.1% SDS for 20 to 25 min at

55°C). The filter was subjected to autoradiography was for 24 hr at -70°C with one intensifying screen. Recombinant plaques were identified, removed as plugs from the agar plate to 1 ml SM solution (0.1 M NaCl/10 mM MgSO₄·7H₂O/50 mM Tris-HCl, pH 7.5/0.01% gelatin), and their titres determined. Each was used to reinfect NM539 cells and plates were incubated overnight to the point of confluence. Phage stocks were collected from the plate lysates following diffusion into overlaid SM solution, and their DNA isolated by the method of Davis et al. (1980). Southern blot analysis under high stringency conditions, again using the 923-bp *Bgl*I repeat probe, confirmed 27 recombinant phage to contain repeat loci.

Ten phage clones were selected and their DNA inserts were released from the phage arms by single or double digestion with 6-bp recognition site restriction endonucleases (*Bam*HI, *Eco*RI, *Hind*III, *Sal*I, or *Sac*I). Restricted fragments shown by Southern blot analysis to contain repeats were isolated from the gel (Geneclean) and subcloned into pTZ18R. The genomic insert from one subclone, pASBgI22.9, was digested to completion with *Alu*I, *Hae*III or *Hpa*II. Fragments were ligated (using T4 ligase, Boehringer-Mannheim) into the replicative forms of M13mp18 and/or M13mp19 and used to transform competent *E. coli* MC1061 cells, followed by cotransfection of JM109 cells. Following isolation (Sambrook et al. 1989), single-stranded DNA was sequenced according to the dideoxy chain-terminating method (Sanger et al. 1977) using the Sequenase Version 2.0 sequencing kit (USB Corp., Cleveland, Ohio) and -40 universal and reverse primers. Sequencing reactions were fractionated on 6% or 8% polyacrylamide (20:1, bis-acrylamide, Bio-Rad)/7 M urea (Sigma) gels using a

SequiGen sequencing apparatus (Bio-Rad) and constant power of 40 W.

3.3.2 Cloning of the *NheI* and *PstI* repeats

The 380-bp *NheI* band of a single salmon individual was isolated from a gel, ligated into *XbaI* cut, phosphatase-treated pUC18 and used to transform competent *E. coli* DH5 α cells. The 570-bp *PstI* band was similarly cloned into *PstI* cut pUC18. Recombinant clones were identified by colony hybridization using Hybond N or Biotyne nylon membranes and, as probes, the isolated gel fragments radioactively labelled as above.

Homologous *NheI* fragments from a brown trout individual and *PstI* fragments from representatives of various Salmonidae species were isolated in the same manner, except that Atlantic salmon clones ASNHE6 (see Section 6.1) and SALT1-SSal28 (see Section 5.1), respectively, were used as probes for their identification. Plasmid DNA was isolated either by protocols of Sambrook et al. (1989) or with the aid of the Magic Minipreps DNA Purification System (Promega). The double-strand DNA was sequenced using T7 polymerase (Pharmacia) or Sequenase (USB Corp.) kits.

3.3.3 Isolation of a complete SALT1 repeat

The λ EMBL3 Atlantic salmon genomic library was screened by plaque hybridization as described in section 3.3.1 using clone SALT1-SSal28 as probe. A single recombinant phage clone (SALT1-SSal1) containing the repeat element was selected and following isolation its DNA was cut with *StuI* and blunt-end ligated into

pUC19. Two plasmid subclones were isolated which permitted sequencing in both directions from a *Stu*I site within the repeat. Primers (designated a, b, and c, Table 3.1) used for sequencing were synthesised with a Milligen oligonucleotide synthesizer.

3.3.4 Isolation and sequencing of the mini- and microsatellite loci

In the course of isolating UV-fluorescent bands of *Bgl*II, *Pst*I, and *Sac*I-digested DNA from agarose gels, a number of clones were found to contain micro- or minisatellite arrays.

The λ EMBL3 genomic DNA library was also screened by plaque lift hybridization as described in Section 3.3.1 using as probe the salmon minisatellite DNA insert of plasmid clone SsPstIL.26 (Section 7.1) or poly(dA-dC)poly(dG-dT) (Pharmacia). Then, in order to isolate and sequence the minisatellite loci identified within the genomic library, DNA from each of the positive phage clones was digested with *Alu*I or *Hae*III, ligated into *Sma*I digested, phosphatase-treated pUC18, and used to transform competent *E. coli* DH5 α cells. Colonies of recombinant bacteria were transferred to gridded plates, and those containing salmon mini/micro-satellite DNA sequence were identified by duplicate colony hybridization using Hybond N nylon membranes and SsPstIL.26 and poly(dA-dC)poly(dG-dT) as probes. Plasmid DNA was sequenced with either T7 polymerase (Pharmacia), Sequenase, or fmol DNA Sequencing System (Promega) kits.

The names of all simple sequence repetitive DNA loci are prefixed with the letters Ss (*S. salar*). Clones derived from fragments purified from ethidium bromide-

stained agarose gels are further identified by restriction enzyme (*Bgl*II, *Pst*I, or *Sac*I), band excised (*Upper*, *Middle* or *Lower*), and clone number (e.g., SsBgHIU.20). Salmon sequences isolated from the genomic library are named by *Alu*I or *Hae*III and include phage clone and plasmid subclone numbers (eg, SsAluI16.70). SsHpall.24 (Section 7.3) is a subclone of pASBgI22.9 (see Section 3.3.1), while the sequence of SsStuI6.62 (Section 7.2) is from the 3' end of a *Stu*I fragment of SALT1-SSaI1 (Section 3.3.3).

3.4 PCR conditions

3.4.1 PCR of rDNA genes associated with *Bgl*I repeats

In order to confirm the proximity of *Bgl*I repeats to rDNA genes, DNA from recombinant λ EMBL3 or pTZ18R clones was included in PCR reactions containing oligonucleotide primers derived from conserved eukaryote rDNA gene sequence (primers are described in White et al. (1990) and Egger and Sigler (1993), and shown in Table 3.1). Reactions were carried out in 25 μ l volumes with 10 pmole of primer in a cocktail containing 10mM Tris-HCl, pH 9.0/50 mM KCl/0.1% Triton X-100/2 mM $MgCl_2$ /0.5 mM each of dATP, dCTP, dGTP, and dTTP (Pharmacia), and 1.5 U Taq DNA polymerase (Promega). A Perkin-Elmer Cetus DNA Thermal Cycler and the following conditions were used: initial denaturation 1 min 15 sec; 30 cycles: 94°C, 1 min; 42°C, 1 min; 72°C, 2 min, increased 1 sec per cycle). PCR products were separated on a 2% NuSeive, 0.5% SeaKem (FMC BioProducts) agarose gel.

TABLE 3.1. PCR primer sequences used in this study

Primer Name		Sequence	Size
A. AMPLIFICATION OF rRNA GENE REGIONS IN THE <i>Bgl</i> II CLONES (Section 4.3)			
NS1	K	GTA GTC ATA TGC TTG TCT C	19mer
NS2	L	GGC TGC TGG CAC CAG ACT TGC	21mer
NS7	T	GAG GCA ATA ACA GGT CTG TGA TGC	24mer
NS8	U	TCC GCA GGT TCA CCT ACG GA	20mer
ITS9mun	A	TGT ACA CAC CGC CCG TCG	18mer
ITS10mun	B	GCT GCG TTC TTC ATC GAT	18mer
ITS1	M	TCC GTA GGT GAA CCT GCG G	19mer
ITS5	N	GGA AGT AAA AGT CGT AAC AAG G	22mer
ITS4	P	TCC TCC GCT TAT TGA TAT GC	20mer
NL5mun	J	GCA TAT CAA TAA GCG GAG GA	20mer
NL10mun	D	GGA ACC TTT CCC CAC TTC	18mer
B. SEQUENCING OF SALT1-SSaII AND INTER-SALT1 PCR (Section 5.4)			
	a	GCT CAA TTT CGA GTC TCA T	19mer
	b	CGG TCA GCT CTA TGA AGA G	19mer
	c	CCC TAA GCA CAC AGC CAA GA	20mer
C. SINGLE-LOCUS VNTR PCR OF THE <i>Ssp</i> StIL.48 LOCUS (Section 7.4)			
	5'	GAC CTT CTG TTT AGT GAA T	19mer
	3'	CAC ATG CGA AGA TCA TCC G	19mer

3.4.2 Inter-SALT1 PCR

Primers a or c (Table 3.1 and Section 5.4), directed outwards from the SALT1 element, were end-labelled with polynucleotide kinase (Pharmacia) and [γ - 32 P]ATP (Amersham) in a buffer prepared according to Sambrook et al. (1989). Primers were used singly in PCR amplifications of approximately 100 ng of genomic DNA and reactions were prepared as described in section 3.4.1. The following amplification conditions were used: initial denaturation 95°C, 5 min, followed by 37 cycles: 92°C, 45 sec; 55°C, 45 sec; 72°C, 90 sec. Sixteen μ l of stop dye (Sequenase, USB Corp.) was added to each reaction. Products were denatured at 70 °C for 5 min and separated on a 6% denaturing polyacrylamide gel at a constant power of 40 W for 4.5 hr. Autoradiography was at -70°C with one intensifying screen.

3.4.3 Single-locus DNA PCR

Primers complementary to sequences flanking the repeats within clone SsPstIL.48 (Table 3.1 and Section 7.4) were synthesized. One primer was 5' end-labelled and reactions were prepared as described in section 3.4.1. PCR conditions were: initial denaturation 95°C, 5 min, followed by 30 cycles: 92°C, 45 sec; 55°C, 45 sec; 72°C, 90 sec. Products were denatured and separated on a 6% polyacrylamide denaturing gel as above.

3.5 Determination of the phylogenetic distribution of repeat elements

For all zoo blot analyses, approximately 7 to 12 μ g of genomic DNA from

selected species was digested with restriction endonuclease, separated in 0.8% to 1.0% agarose gels and transferred to Pall Biodyne nylon membranes. Hybridization with the appropriate radio-labelled probe was at 37°C and otherwise as described in section 3.3.1, followed by washing at low stringency (twice in 2× SSC for 20 min at room temperature; twice in 2× SSC/0.1% SDS, twice in 1× SSC/0.1% SDS for 20-25 min each time at 50°C). Autoradiography was at -70°C with one intensifying screen for varying lengths of time.

In the case of the salmon *Bgl*II repeat, hybridization studies were carried out using the insert from pASBgl22 subclone mpAlu18 as the probe. Detection of the taxonomic distribution of the *Tc*I-like element employed, variously, inserts from clones SALT1-SSal9, -SSal28, and -SSal44. In determining the intraspecific variation of the *Nhe*I repeat, Atlantic salmon clone ASNHE6 and a brown trout locus, BTNHE68 (94% identical to ASNHE6, Section 6.5), were employed as probes.

In the case of the mini-satellite sequences, plasmid inserts from clones noted in Section 7.5 were used as probes, with the exception of SsSacIU.44. This clone due to mismanagement was lost, so a 37-mer oligonucleotide probe was synthesized (5'-GAGCTGACAACAGTCAGTCAGTGTTTCCCTACTATTA-3') and end-labelled with [γ -³²P]ATP as described in Section 3.4.1.

3.6 Fingerprint analyses

In preparation for DNA fingerprint analyses using minisatellite probes SsBglIIU.20 and SsPstIL.26 (Section 7.6), approximately 4 µg of genomic DNA from

Atlantic salmon of populations in Canada, Ireland, and Scotland were digested with either *PvuII* (an isoschizomer of *HaeIII*) or *AluI*, ethanol precipitated, redigested a second time with the same enzyme, and subjected to electrophoresis in 1.0% agarose gels. A modified DNA hybridization protocol originally devised at the Central Forensic Laboratory of the Royal Canadian Mounted Police, Ottawa (R. Fournery, pers. comm.) was used. Briefly, nylon membranes were prehybridized 2 to 5 hr and hybridized with the radiolabelled DNA insert of clone SsBgIIIU.20 or SsPstIL.26 (Fig. 1A,E) for 20 hr in 1.5× SSPE (150 mM NaCl/8.7 mM NaH₂PO₄-H₂O/1mM Na₂EDTA-2H₂O, pH 8.0)/10% polyethylene glycol/7.0% SDS at 60°C, followed by high stringency washing (twice in 2× SSC for 10-20 min at room temperature, followed by twice in 2× SSC/0.5% SDS, twice in 0.2× SSC/0.5% SDS, and twice in 0.1× SSC/0.1% SDS at 60°C for 20-25 min each time). The blots were then briefly rinsed twice in 2× SSC at room temperature. Hybridizations for all of the DNA fingerprinting and some of the zoo blot analyses were performed in a Robbin's Scientific Model 400 Hybridization Incubator. Washing of the nylon membranes was performed either directly in the roller bottles or in pyrex dishes in a heated shaking water bath.

3.7 DNA sequence analysis

The relationships of the ASNHE and BTNHE repeat sequences and of the salmonid SALT1 homologues were examined with the aid of PHYLIP (Phylogenetic Inference Package, ver. 3.4, Felsenstein 1991). One hundred multiple data sets were generated by the program SEQBOOT, and distance measures were then calculated by

the program DNADIST under the Kimura 2-parameter model (Kimura 1980). Distance matrices were analyzed by nearest neighbour-joining (program NEIGHBOR) or Fitch-Margoliash and least squares (program FITCH) methods. For each analysis unrooted majority-rule consensus trees were constructed using the CONSENSE program.

Nucleotide sequences were compared with the NCBI/NIH GenBank database using the BLAST E-mail server (Altschul et al. 1990), and the EMBL database using FASTA (the FastA sequence search server) (Pearson and Lipman 1988). The translated amino acid sequence of the predicted transposase region of SALT1-SSa11 was compared with the SWISS-PROT and PIR protein database using BLITZ, ver. 1.4 (MPsrch protein sequence search server) (Sturrock, S.S and Collins, J.F., Biocomputing Research Unit, University of Edinburgh) and FASTA. Original sequences were annotated using the Authorin program (Intelligenetics, Inc.) and submitted to the GenBank database.

The editing and alignment of sequences were performed with the aid of ESEE, ver 2.0 (Cabot and Beckenbach 1989). The programs AllAll of the Darwin server (Computational Biochemistry Research Group, Swiss Federal Institute of Technology, E.T.H., Zurich) (Gonnet et al. 1992) and MultAlin ver 3.0 (Corpet 1988) were useful in optimizing sequence alignments. Programs available in the PCGENE (Intelligenetics, Inc.) package also were useful for aligning sequences and searching for repeats and restriction enzyme recognition sites.

3.8 Determination of repeat copy number

Three methods were used to determine repeat copy number:

1) To quantify the relative amount of DNA represented by the ethidium bromide staining bands on agarose gels, 35mm polaroid negatives of the gel were scanned with a LKB 2222-020 UltraScan XL densitometer. In addition to the lane containing digested DNA, a neighbouring lane containing no DNA was scanned to provide a measure of background signal. With the aid of the GelScan XL software package (Pharmacia), this background was subtracted from the DNA signal, and the total integrated area under the peak representing the bright band was then calculated. That area of the peak above a tangent drawn to the scan plot was assumed to represent repetitive DNA. Limitations of the software required that this area be estimated by cutting the region out of a paper plot and, using a Mettler AE166 analytical balance, estimating the area under the peak to the tangent as a weight percentage of the total area under the peak to the background baseline. It was assumed in calculating copy number that the 'c' value for the salmonid haploid genome is 3.0 pg (Hinegardner and Rosen 1972).

2) When the λ EMBL3 genomic library was screened with a repeat probe, the number of positive plaques as a percentage of the total number of recombinant clones was calculated. Knowing the sequence length of a single repeat unit, and assuming the average length of a phage insert to be 15 kb, it was then possible to estimate the number of unit repeat copies in the entire genome.

3) Dot blot hybridization has been used in other studies (Franck and Wright 1993; He et al. 1992; Elder and Turner 1994; Garrido-Ramos et al. 1994) as a means of estimating repeat copy number in fish. Atlantic salmon genomic DNA samples of

known concentrations were serially diluted 1:10 each time. Known concentrations of purified repeat DNA (the *Bgl*I 923-bp band, pASBgl22 subclone mpAlu18, SALT1-SSal28, and SALT1-SSal44) were diluted in a similar manner. All DNA was denatured for 15 min at room temperature in 2.0 M NaOH/20 mM EDTA (and in the case of SALT1-SSal44 only, neutralized with an equivalent volume of ice-cold 2M NH₄Ac) and placed on ice. Ten to 25 μ l volumes of each aliquot were slowly pipetted onto a nylon membrane using a vacuum dot blot apparatus (constructed by Technical Services, Memorial University). The membrane was rinsed in 2 \times SSC and baked at 80°C for at least one hour. Probing was at high stringency using as probes inserts from the repeat clones. Several problems were encountered including poor reproducibility, uneven hybridization of the probe to the dotted DNA and, in some cases, variation in hybridization intensity between sequential dilutions which did not appear to correspond to the actual declining DNA concentrations. For these reasons, reliability of the dot blotting results was in doubt and they were not considered when quantifying repeat copy numbers.

3.9 Cytogenetic techniques

3.9.1 Chromosome preparations directly from embryos

Chromosome preparations were made directly from whole salmon embryos during the second to fourth week post-fertilization by a modification of the "chopping method" of Yamazaki et al. (1981). Briefly, whole eggs were immersed in water containing 0.1 mg/ml colchicine (Fisher Scientific) or 0.1 μ g/ml colcemid (Gibco/BRL)

for periods of time ranging from 2 to 8 hr. The egg chorion was then removed, and the embryo extracted in saline solution (0.134 M NaCl/3.8 mM KCl/3.2 mM CaCl(2H₂O)) and immersed in hypotonic solution (0.075 M KCl or 0.8% Na-citrate) for 20 to 25 min. The embryo was then transferred to Carnoy fixative solution (75% methanol/25% acetic acid) for 15 min at 4°C, followed by three changes of Carnoy for 30 min each at -20°C. A single whole embryo was transferred to a drop of water on a slide, chopped to a paste of cells with the blade of a scalpel, and flooded with several drops of Carnoy solution. The slide was then flamed over a Bunsen burner and dried on a slide warmer set to 42°C.

3.9.2 Chromosome preparations from tissue cultures

Tissue cultures were prepared from liver or whole embryos. The tissue was dissected in saline solution and minced with a sterile scalpel. Tissue was cultured in sealed Ambitubes (Fisher) in 2 ml of Minimum Essential Medium (MEM) with Hank's salts and L-glutamine (Gibco\BRL) or RPMI Medium 1640 with 25mM HEPES buffer and L-glutamine (Gibco\BRL). Medium was augmented with 16% Fetal Calf serum (Gibco\BRL) and 80 µg/ml Gentamycin. To some tubes 1 to 5% of the mitogen phytohemagglutinin (PHA-M, Gibco\Brl) was added. Medium was changed once a week and prior to confluence the cells were harvested either for chromosome preparation or subculturing.

Prior to harvesting of cells, 0.1 µg/ml colcemid was added to each tube and allowed to sit for 5 to 8 hr. One ml of 1× Trypsin-EDTA (Gibco\BRL) was then

dispensed to each 2 ml of culture media and allowed to sit with occasional shaking for 1 min. This solution was discarded and replaced by a further 1 ml of Trypsin-EDTA solution. The tube was gently shaken for 5 to 10 min until cells had dislodged from the wall of the culture flask. Five ml of hypotonic solution were then added and allowed to sit for 20 min, followed by centrifugation at 1000 rpm for 5 min and fixing with three changes of cold Carnoy solution. The cell suspension was dropped on to a prewarmed slide and dried on a slide warmer.

3.9.3 Chromosome preparations from blood

Blood was collected from the caudal blood vessels of adult Atlantic salmon and transferred to 3 ml vacutainers containing Na-heparin (Becton Dickson). Either whole blood or enriched leukocytes were used for short-term blood culture. Leukocytes were collected in a buffy layer by centrifugation twice at 350 rpm and once at 800 rpm for 10 min at room temperature. White cells and plasma were removed each time by pasteur pipette, pooled, and 0.5 ml was inoculated into 5 ml of MEM, RPMI or 199 (Gibco/BRL) medium, prepared as above. In addition to adding 5% PHA-M to some cultures, effects on cell proliferation of pokeweed mitogen and lipopolysaccharide (Sigma) were also tested.

Cells were harvested following 5 to 14 days incubation at 37°C and normal atmosphere. At least 5 hr prior to harvesting, 0.1 µg/ml colcemid was added to each tube. Cells were concentrated by centrifugation and subjected to hypotonic treatment and fixation as described in Section 3.9.2.

3.9.4 Chromosome banding

To induce restriction enzyme banding, slides were pretreated with 15 to 30 units of *AluI*, *BglI*, *HaeIII*, *HpaII* or *MboI* in 30 μ l of restriction buffer for at least 5 hr and stained with 0.3% Wright's stain (1:4 dilution with Gurr buffer, pH 6.8) for 60 sec. For G-banding the slide was heated at 90°C for 30 min, and treated with 0.008% trypsin in Hank's buffered saline solution, pH 7.4 (Gibco\BRL) for 45 to 90 sec, followed by washing twice in 0.9% saline solution. Staining was with 2% Giemsa for 45 sec to 5 min.

3.10 Pulsed field gel electrophoresis (PFGE)

High molecular weight DNA was isolated in agarose plugs by the following procedure. Ten μ l of freshly drawn blood cells were washed twice in Hank's buffered saline, resuspended to a final volume of 200 μ l, and counted using a haemocytometer. Blood cells diluted to 8×10^7 cells/ml were mixed with an equal volume of 1.6% low gelling temperature agarose (Sigma Type VII) and pipetted into 10 cm lengths of silicon tubing (I.D. 0.238 cm; Cole-Parmer Instrument Co., Chicago). When hardened, the column of agarose was blown from the tubing and cut into lengths suitable for insertion into individual wells of an agarose gel. These were placed in a Falcon tube containing 10 ml lysis buffer (250 mM EDTA/55 mM Tris-HCl/1.1% Sarkosyl) made 0.25 mg/ml with proteinase K (Sigma). Following overnight incubation at 50°C, the lysis/protease K solution was replaced and the plugs were left for 2 more days at 50°C. Plugs were stored in lysis buffer at 4°C (Sambrook et al. 1989).

Prior to restriction enzyme digestion, all plugs were washed twice in 1 mM phenylmethylsulfonyl fluoride (Sigma) for at least 2 hr each time, and then 5 times in TE buffer. Whole plugs were digested overnight in 100 μ l of restriction buffer containing 500 μ g/ml nuclease-free BSA (Sigma) with 30 units of enzyme. Sample DNA plugs, together with plugs containing *Saccharomyces cerevisiae* whole chromosomal DNA (Bio-Rad), were pushed into wells and covered with a drop of molten agarose. Electrophoresis was performed in circulating 0.5 \times TBE buffer chilled to between 14°C and 17°C, using a CHEF-DR II pulsed field electrophoresis system with Pulsewave 760 Switcher (Bio-Rad). Various ramping times and pulse rates were tested. DNA in PFGE gels was depurinated in 0.25 M HCl for 20 min prior to preparation for Southern blot transfer as described in Section 3.2.

CHAPTER 4. THE *BGLI* REPEAT - RESULTS AND DISCUSSION

4.1. Cloning of the repeat sequence

The Atlantic salmon genome was screened for the presence of repetitive elements by digestion of nuclear DNA with over 40 restriction enzymes, followed by electrophoresis of the DNA through agarose gels and ethidium bromide-staining. In addition to the tandemly arrayed *HindIII* element previously identified by Moir (1988), *AluI*, *AvaII*, *BglI*, *BglII*, *BstEII*, *EcoRV*, *HaeIII*, *HpaI*, *HpaII*, *NheI*, *PstI* and *SacI* digestion revealed repetitive DNA in the Atlantic salmon genome (Fig. 4.1). In this chapter, I describe a previously uncharacterized, highly repetitive element existing in tandem arrays and identified by the restriction enzyme *BglI*.

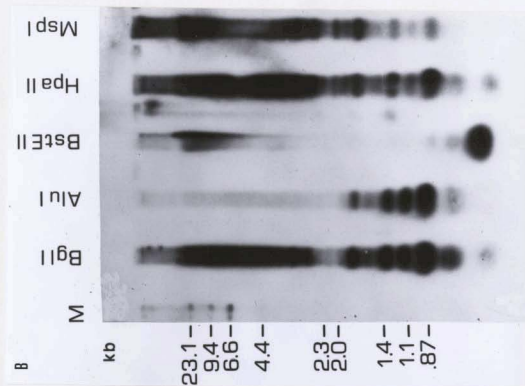
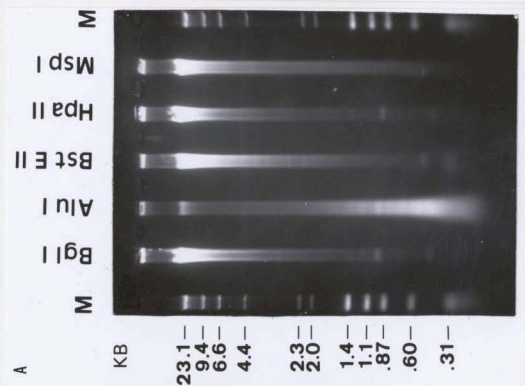
In the case of *BglI*, two bright bands were visible upon ethidium bromide-staining: one 923 bp and the other more diffuse but centred at 430 bp (shown in Fig. 4.1, and more clearly in Fig. 4.2A). When the 923-bp *BglI* band was isolated and used as a probe, cross hybridization was detected to bright bands visible upon restriction digestion of DNA with *AluI*, *HpaII*, and *BstEII* bands (Fig. 4.2B). It is notable that when the restriction enzyme *MspI* was used, the two bands visible with *HpaII* virtually disappeared (Fig. 4.2A), and when hybridized to the 923-bp DNA band yielded a fainter signal. *MspI*, an isochizomer of *HpaII*, also recognizes the sequence CCGG but is not inhibited by methylation of the second cytosine residue. It is not clear how methylation might yield this digestion pattern, and the phenomenon was not investigated further.

The radiolabelled 923-bp *BglI* band was also used to screen an Atlantic salmon DNA genomic library cloned in λ EMBL3. Following autoradiography for 24 hr, 43

Fig. 4.1. Ethidium bromide stained 1.2% agarose gel of Atlantic salmon DNA digested with restriction endonucleases that produce bands over the background smear. Such bands indicate the presence of repetitive elements. The size markers in kb pairs are derived from the migration of λ DNA cut with *Hind*III and ϕ X174 DNA cut with *Hae*III.



Fig. 4.2. A, left) Atlantic salmon DNA digested with *Bgl*I, *Alu*I, *Bst*EII, *Hpa*II, and *Msp*I, and electrophoresed in a 0.8% agarose gel. Sizes of the markers shown at left are λ DNA cut with *Hind*III and ϕ X174 DNA cut with *Hae*III. B, right) Southern blot (Hybond N) analysis of gel in A), probed with the 923-bp *Bgl*I band isolated from an agarose gel. Hybridization was at 42°C as described in section 3.3.1, and washing was at high stringency (twice in 2 \times SSC for 25 min at room temperature, 4 times in 0.1 \times SSC/0.1% SDS for approximately 30 min each time at 60°C).



strongly positive plaques were detected out of a total of 7000 screened. DNA from stocks of recombinant phage was isolated, and 27 inserts were confirmed by Southern blot analysis to contain salmon DNA inserts with the *Bgl*I repeat. Using 6-bp recognition site restriction endonucleases, DNA was trimmed from the flanking regions of selected phage inserts, and the shortened fragments were subcloned into pTZ18R. The shortest plasmid insert, pASBg122, was 5.3 kb. Restriction mapping (Fig. 4.3), partial digestion of the insert with *Bgl*I (Fig. 4.4), and sequencing (see below) indicated that the repeat elements occupied about 2.9 kb of the subcloned DNA. These repeats were arranged in tandem, although as indicated by partial digestion of the plasmid insert, multimer as well as monomer repeat units were present (Fig. 4.4). The insert of pASBg122.9 was digested with *Alu*I, *Hae*III, or *Hpa*I and "shotgun" subcloned into M13 vector for the purpose of sequencing its *Bgl*I repeat units.

4.2 Sequencing of pASBg122

Sequences of the pASBg122.9 *Bgl*I repeats, beginning at their *Alu*I sites, are shown in Fig. 4.5. Their tandem nature is confirmed by several *Hae*III and *Hpa*I subclones which traverse *Alu*I sites and continue into adjoining repeats. Four variant sequences were found among 18 subcloned repeats. Variant-2 is 403 bp in length, consisting of two related dimerized subunits of 212 and 191 bp. The recognition sites for *Bgl*I and *Hpa*I are located in a unique G+C-rich region near the beginning of the second subunit. Variant-2 corresponds to the 430-bp band visible in restriction enzyme-digested DNA, minus a deleted 26-bp stretch between nt 189 and 215. These 26 nt are

Fig. 4.3. Map of subclone pASBgl22.9, showing the location of the *Bgl*I repeats. A 5.3 kb fragment from phage clone ASBgl22 of the genomic library was subcloned into the *Sac*I and *Hind*III sites of the plasmid cloning vector pTZ18R. Assuming that the average *Bgl*I repeat unit is less than 430 bp, 7 would be expected to reside within the 2.9 kb repeat array. Restriction enzyme recognition sites in the 5' region were determined by sequencing from the plasmid cloning site (see Fig. 4.5B). Distance to the 5' start point of the repeat array is uncertain. Restriction sites in the 3' flanking region were determined by sequencing (Fig. 4.5C) and restriction mapping by single- and double-digests of the plasmid insert.

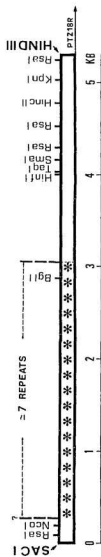


Fig. 4.4. Southern blot analysis of *Bgl*I partial digests of the salmon DNA insert of subclone pAS*Bgl*22.9. Recombinant DNA was electrophoresed through a 0.8% agarose gel, blotted to Hybond-N nylon membrane, and probed with the genomic insert of M13 subclone mp18Alu18 (see Fig. 4.5A). Hybridization was performed under high stringency conditions (see section 3.3.1). Lanes from left to right: 0, 9.0, 4.5, 2.2, 1.1, 0.56, 0.28, 0.14, 0.07, 0.035, and 0.018 units of *Bgl*I. Aliquots in lanes 3 to 11 were digested for 1 hr at 37°C. Those in lanes 1 and 2 were digested for 9 h. Sizes of markers shown at left are *Hind*III digested λ DNA.

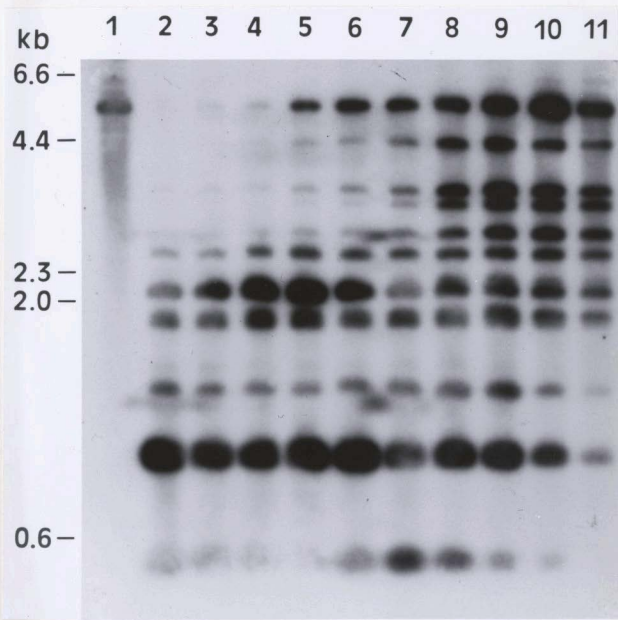


Fig. 4.5. A) Sequence of *Bgl*I repeat variants in pASBg122.9, obtained from M13mp18 or M13mp19 subclones. Some *Alu*I fragments were cloned in reverse orientation (rev). Sequences of the three variants were obtained from the following subclones: Variants 1) mp18Alu1, mp18/mp19Alu6, mp18Alu20, mp18/mp19Alu14, mp18/mp19Alu7, mp18Hae3(rev), mp18Hae19; 2) mp18Alu3(rev), mp18Alu4, mp18Alu16, mp18Alu8(rev), mp18Alu13(rev), mp18Alu23, mp18Hpa24; 3) mp18Hpa20, mp19Hpa9, mp18Hae13. Stretches of repeated nt within the two subunits of Variant-1 are underlined twice (==). Those repeated within Variant-2 are underlined once (—). Variant-4 (not shown) is the same as Variant-1, with the exception of a deletion of nt 190 to 215. Variant-3 was isolated as a *Hpa*II subclone commencing at nt 263; after 100 bp it degenerates into flanking sequence unrelated to the repeat (16 nt of this are shown in lower case). Identical nt are defined by ".", missing nt by "-". A portion of the *Bgl*I repeat is also compared with the sequence from human *Hinf*I family repeat DNA (HUMRShinf; Shimizu et al., 1983, GenBank Acc. #K00580). Identity is 65% in a 122-nt overlap. B) 5' flanking region of pASBg122.9. Double-stranded plasmid sequencing was accomplished using the M13 reverse primer. Sequence between the two *Alu*I sites was obtained from a partially overlapping M13 subclone (mp18Alu22). C) 3' flanking region of pASBg122.9. The M13 universal primer was used.

A.

```

      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
variant AluI                                     HaeIII DdeI                                     DdeI
1 AGCTATATAATTCCTCTTGCTGTATCTCATTTAGTTGGGCCCTTAGGATGGTGCTCTAA
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 GGGTGGGTATGGAGGTCCCTATACCTCTATAAGTATTTAAAAACCGTTTTTCAAAAATGG
2 .....T.....
3 .....#3 continued from nt 345.....A.....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 TTGTGGTAACCCAAA-----ACATAGTGCAATTTCTGAGAAGA
2 .....TAAACACAGGTGCCCGTTGG..T.....
3 ..CC.....TAAACACAGGTGCCCGTTGG..T...ctctgaagcgaactga
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 AAACATAGCCGATTTTCTGAGAAGAAAAACATAGAAATATTTTCTGAAGATTTTTCAAA
2 .....
HUMRS .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 CT-TCCATAATCAGCCA-GGCCGGCCCCCATTTAGCTTGGGCCCGTAGTAGGGTGCTGTCT
2 .....
3 .....start of variant #3.....
HUMRS G.G.....CT..G..AG.....G..TACA..G.....A...CCT
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 AGGGTGGGTATGGAGGTCCCTATACCTTTAAAGTATTTAAACAGTTATAAAATTCA
2 .....
3 .....#3 continued at nt 105.....
HUMRS .TA....C..TA.TA.GAGA..GA..GAA..GT.C....GG..CA.AT..T....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 CCCCTGGTAACCCAAATGCTACGGACACGTCCCCACCAAACATAGAATATTTCTGAGA
2 .....AG
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 AGAAAAAACATAGCCGATTTTCTGAGAAGAAAAACATAGAATATTTTCAAGAGGATTTT
2 .TTTTTTGAAA (AGCT)
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 AAAAATCTCCATAAAACACTCATTCTTCATCCACTGACATGGGACAGCAGTGGGGTGCT
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 GAAACAGGTGTCCTGGGGTGGTGATGGCCTTTAAAGTAATTTAGAGCCTTTTAAAAAT
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 GTCACCCCTGGTAAGCCAAATGCTACGGACACGTCCCCACCAAACATAGTGCAATTTCT
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 GAGAAGAAAAACATAGAATATTTTCAAGGATTTTAAAACTTCCATAAAACACTCA
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 TTCTTCATCCCACTGACTTGGGACCAGTGGGGTGCTGAAACAGGTGTCCTGGGGTGG
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 TGATGGCCTTTAAAGTATTTAAACAGTTGTAATTTTCAACCCCTGGTAACCCAAATGC
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 TACGGACACGTCCCCACCAAACATAGTGCAATTTCTGAGAAGAAAAACATAGCCGATT
2 .....
      10      20      30      40      50      60
      |-----|-----|-----|-----|-----|
1 TTCTGACAAGAAAAACATAGTCATTTTCTGAAGATTTTITGAAA (AGCT)

```

B.

10 20 30 40 50 60
 -----|-----|-----|-----|-----|
 SacI MboI
 GAGCTCAATATCCTCCACCAAGCTGGGAAAAATCGGATCAATACTTAGGACCAATCG
 RsaI
 GGCGCCAGAGGGATATGGTCCAAGTACCCACGGTTCTGAGGGGCTTGTAACCTCAAAGT
 AluI
 GTCGTAAGCTCTCCCTATTCCGCATGAGTAGTAATCATCAATAGGTCGCCGAGGGGAT
 NcoI
 ATGGACCGAACTACCATGGTTCCGAGGGACAAGCAACCCCTCAGAAGTGTCCGTTTCGCAT
 AluI
 TTATTTAATCGGACTAGTAGTTTGCCACTATCCCCGAGCT

C.

HpaII AluI HaeIII
 CCTCCTGTGGTCCTGCTGCCCCTCTCTCTGCGGCTCCGGACTCTAGCTCCTGGCCCT
 RsaI
 TCTCTGTACACCCTGGTAACCCGTTATAAAATAAGGGGGGAGGTGGAGGAAGACGCCTTC
 HindIII
 CGTCCCGACAAAAGCTT

present in two tandem copies in Variant-1, which consequently possesses extra *DdeI* and *MboI* restriction recognition sites. A duplication event such as this may be the result of slipped-strand mispairing during a round of DNA replication. One T to A transversion between variants 2 and 1 has also resulted in the loss of a *DdeI* site at position 160.

Variant-1 and Variant-4 (the latter not shown, but identical to Variant-1 without the 26 bp duplication) are 923 and 897 bp, respectively, A+T-rich (61%), and may have arisen from a partial duplication of the first 403 bp unit with extensive loss of homology, including the sites for *AluI*, *BglI* and *HpaII*. Variant-3 has resulted from the fusion of portions of two repeats and is degenerate at its 3' end. The nucleotide sequence of Variant-1 has been submitted to GenBank and assigned accession number L01505.

4.3 Genomic organization of the repeat

To estimate the prevalence of the *BglI* repeat element in the salmon genome, *BglI*-digested DNA was electrophoresed, and photographic negatives of the ethidium bromide-stained gel (Fig. 4.2A) were scanned by densitometry. The integrated areas under the peaks for the 430-bp and 923-bp fragments represented 1.2% and 1.1%, respectively, of the total DNA. Since 'c' values for salmonid fish are 3.0 pg (Hinegardner and Rosen 1972), the haploid genome contains approximately 8.1×10^4 and 3.4×10^4 copies of the two repeat lengths.

Plaque lift membranes, used originally to isolate the *BglI* repeat from the

λ EMBL3 library, were reprobed with a plasmid (pDM238) insert spanning the 18S and 28S regions of the rDNA repeat of *Drosophila melanogaster* (gift of J. Hancock, Cambridge University, U.K.). Under low stringency hybridization conditions, one-quarter of the plaques positive for the *Bgl*I element were also positive for the rDNA gene probe. This suggested that some *Bgl*I repeats are near rDNA genes. Fig. 4.6B shows unrestricted DNA from ten phage clones hybridized with the insert of pDM238. Some signal may be non-specific hybridization due to the large amounts of electrophoresed phage DNA. However, by further screening of recombinant phage fragments known to contain *Bgl*I repeat loci (see section 3.3.1), unequivocal identification of at least two with adjoining rDNA related sequences was possible (Fig 4.6D).

To test this observation, DNA from the 27 recombinant phage isolated from the salmon library was screened by PCR analysis using three sets of primers derived from conserved eukaryote rDNA sequences (White et al. 1990; Egger and Sigler 1993) and designed to amplify portions of (1) the 5' end of the 18S subunit; (2) the 3' end of the 18S subunit, across the internal transcribed spacer region to the 5' end of the 28S subunit; and (3) an internal portion of the 28S subunit (Table 3.1). Of ten successfully amplified phages, 9, 6 and 1 were identified by the three primer sets (Fig. 4.7), respectively. The primers did not amplify host *E. coli* or λ EMBL3 DNA (not shown). It would appear that the *Bgl*I element is upstream of the 18S rDNA subunit, probably in the intergenic spacer (IGS) region. Recently, using two-colour FISH labelling, a DNA fragment obtained by PCR amplification of the *Bgl*I repeat sequence identified

Fig. 4.6. Hybridization of the insert of pDM238 with *Bgl*I repeat phage clones (ASBgl series, as marked on the figures). Plasmid pDM238 contains portions of the 18S and 28S regions of the rDNA repeat of *D. melanogaster*. Hybridization was at 37°C and washing was twice in 2× SSC for 25 min at room temperature, 2 times in 2× SSC/0.1% SDS, and 2 times in 1× SSC/0.1% SDS for 30 min each time at 50°C. A) Intact phage DNA from clones identified as containing *Bgl*I repeats, electrophoresed in a 0.8% agarose gel. Phage clone numbers are shown. B) Southern transfer of gel in A) hybridized with the pDM238 insert. C) Fragments excised from the phage DNA of A) and separated in a 0.7% agarose gel. In the case of clones 2, 4 and 8, multiple fragments were isolated and pooled. (Fragments were identified as containing *Bgl*I repeats by digesting phage DNA with 6-bp recognition site restriction endonucleases and, following electrophoresis and Southern transfer, probing the fragments under high stringency conditions with the 923-bp *Bgl*I band isolated from an agarose gel). D) Southern transfer of gel in C) hybridized with the pDM238 insert.

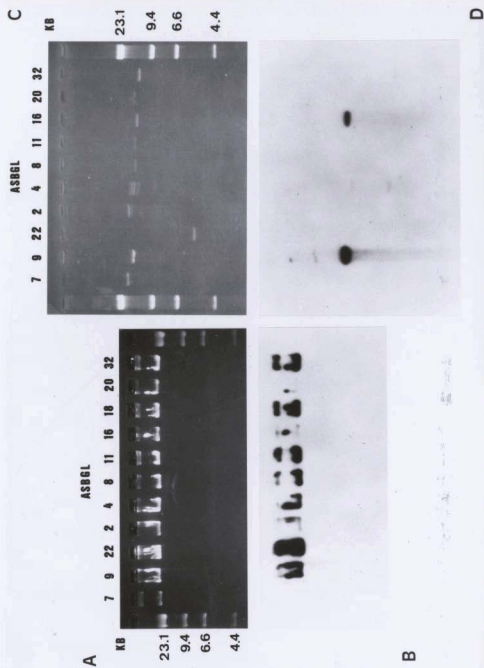
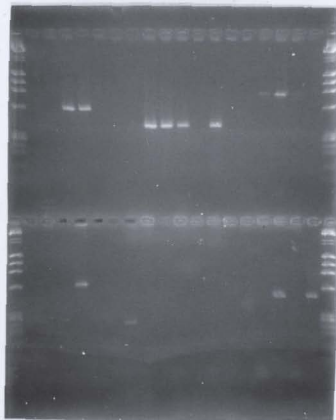


Fig. 4.7. Amplification of selected recombinant λ EMBL3 phage DNA (ASBgl11, 16, 18, 20, 22) and genomic DNA (from LaHave R., Nova Scotia, sample C9) by primers specific to rDNA cistron sequence conserved among eukaryote species. Although all phage samples were selected from among those which had previously yielded PCR products, not all have been successfully amplified here. Control samples (amplification reactions without DNA) are marked as C. Primers are described in section 4.3 and shown in Table 3.1. A Perkin-Elmer Cetus DNA Thermal Cycler and the following conditions were used: initial denaturation 1 min 15 sec; 30 cycles: 94°C, 1 min; 42°C, 1 min; 72°C, 2 min, increased 1 sec per cycle). PCR products were separated on a 2% NuSeive, 0.5% SeaKem agarose gel.

PRIMER K-L T-U A-B
 18S 18S 18S-5.8S

-C-11-C9 -C-11-18-C9 -C-16-20-C9



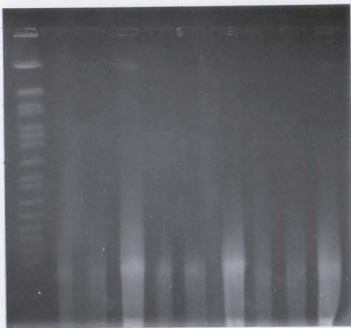
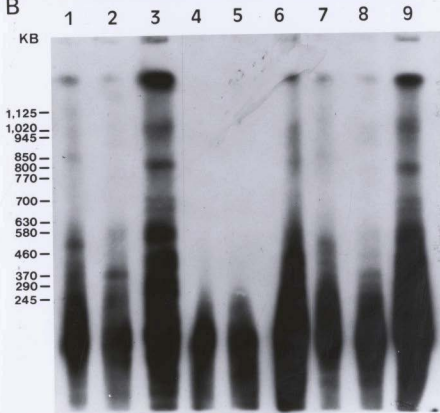
C-18-20-C9 C-22-C9 C-11-18-C9 C-11-20-C9
 18S-5.8S 18S-28S 28S
 M-B N-P M-P J-D

by the present study was confirmed to cohybridize with a rDNA probe to the NOR-bearing pair of chromosomes in Atlantic salmon (Pendás et al. 1994c).

Evidence from pulsed-field gel electrophoresis also offers insight into the genomic distribution of the *Bgl*I element. Agarose-embedded salmon DNA was fragmented with the enzymes *Nhe*I, *Kpn*I and *Sac*I and separated by PFGE. Following Southern blot analysis with the insert of mpAlu18 (Fig. 4.8B), the presence of very high molecular weight bands (> 800 kb) predicts a higher-order periodic organization of some *Bgl*I repeats. Bands were not apparent on the ethidium bromide-stained gel (Fig. 4.8A). RFLP variation between individuals is also seen, most obviously for the *Nhe*I digests, but this may be an artifact caused by nonconsistent amounts of DNA. The phenomenon of subrepeats present within rDNA spacer regions occurring as independent satellite-like DNA elsewhere in the genome has been shown for other organisms (Schmidt 1984; Unfried et al. 1991). In the case of humans, ten percent or more of the IGS regions of rDNA is occupied by *Alu* SINE sequences, and the ribosomal gene complexes themselves sit within large tracts of satellite DNA (Kaplan et al. 1993).

Phillips and Pleyte (1991) and Zhou et al. (1994) report several salmonid species showing both inter-individual and inter-population differences in the lengths of spacer regions of different rDNA repeat units. They suggest that such polymorphisms may prove useful in differentiating fish stocks. RFLPs of the rDNA complex have been used to distinguish populations of mice (Suzuki et al. 1986) and *Drosophila* (Williams et al. 1985). Cutler et al. (1991) detected an RFLP near 18S subunit genes capable of

Fig. 4.8. PFGE analysis of Newfoundland Atlantic salmon DNA. Electrophoretic conditions were 43.2 hr at a pulse interval of 150 sec, followed by 22.5 hr at an interval of 250 sec, with a constant temperature of 14°C and current of 105 V. A) Ethidium bromide-stained agarose gel. B) Southern transfer of gel in A) hybridized with the insert of *Bgl*I subclone mp18Alu18. Hybridization was at 42°C as described in section 3.3.1, and washing was at high stringency at 55°C. Restriction enzymes used for DNA digestion were: lanes 1 to 3, *Nhe*I; lanes 4 to 6, *Kpn*I; lanes 7 to 9, *Sac*I. Atlantic salmon were from: lanes, 1, 4, and 6, N.E. Placentia; lanes 2, 5, and 7, Exploits R.; lanes 3, 6, and 8, Gambo Pond (nonanadromous variety). Sizes of markers shown at left are those *S. cerevisiae* whole chromosomal DNA.

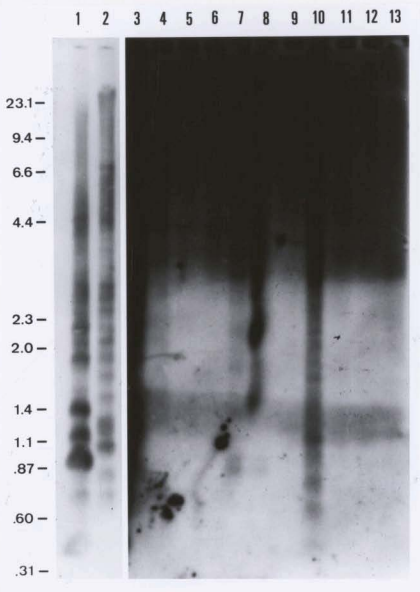
A**B**

distinguishing Atlantic salmon from Newfoundland and Scotland. It is not known if this polymorphism is related to the presence of the *Bgl*II repeat.

4.4 Species distribution of the *Bgl*II repeat

In order to ascertain the phylogenetic distribution of the *Bgl*II repeat element, DNA from Salmonidae fish species other than Atlantic salmon (brown trout; Arctic charr; Japanese white-spotted charr, *S. leucomaenis*; brook trout, *S. fontinalis*; rainbow trout; chum salmon, *Oncorhynchus keta*; chinook salmon, *O. tshawytscha*; pink salmon, *O. gorbuscha*; coho salmon, *O. kisutch*; masu salmon, *O. masou*; Arctic grayling, *Thymallus arcticus*) and a non-salmonid species (northern pike, *Esox lucius*) were digested with *Hpa*II, electrophoresed, and Southern blots were hybridized with the ³²P-labelled insert of a Variant-4 subclone. When low stringency conditions were used, the repeat element was strongly represented in the genomes of Atlantic salmon and brown trout following autoradiographic exposure of 24 hr (Fig. 4.9, lanes 1 and 2), but showed no hybridization with DNA of the other species. After exposure for 9 days (lanes 3 to 13), weak signals were detectable, mainly as non-specific background, but with very faint bands apparent for Arctic charr, Japanese charr, and pink salmon. The *Bgl*II repeat element, if present in species other than those of the genus *Salmo*, retains sequence identity too low for positive detection by Southern hybridization techniques. However, Hartley and Davidson (1994) isolated an Arctic charr repeat fragment (GenBank Acc. #L00991) approximately 180 bp long and flanked by *Dra*I sites, and now shown to be a homologue to a portion of the salmon *Bgl*II repeat (Fig. 4.10). The presence of *Bgl*II

Fig. 4.9. Zoo blot analysis of the distribution of *Bgl*II elements in Salmoninae species, Arctic grayling, and northern pike. Approximately 7 to 12 ug of DNA from an individual of each species was digested with *Hpa*II and separated in a 0.8% agarose gel. Following Southern blotting, hybridization was to the insert of M13 subclone mp13Alu18 (see Fig. 4.5) at 37°C under low stringency conditions (see section 3.5). Lanes 1 to 2, 24 hr autoradiographic exposure; lanes 3 to 13, 9 day exposure. Lanes 1, Atlantic salmon; 2, brown trout; 3, Arctic charr; 4, rainbow trout; 5, chum salmon; 6, chinook salmon; 7, pink salmon; 8, coho salmon; 9, masu salmon; 10, white-spotted (Japanese) charr; 11, brook trout; 12, Arctic grayling; 13, northern pike.



repeats in Pacific salmon is unconfirmed. Based on limited fossil evidence, *Oncorhynchus* is believed to date from at least 6 million years ago (Ma) (Stearley and Smith 1993; Nelson 1994), so it is possible that amplification of the *Bgl*I element to significant copy numbers in salmonines occurred in more recent times.

The 923-bp *Bgl*I sequence was submitted to GenBank. Although no sequence of obvious homology was found, the highest score as determined by the FASTA search algorithm (opt score 169) belonged to a human satellite DNA repeat element (Shimizu et al. 1983, Fig. 4.5A). The human *Hinf*I family occurs in tandem arrays of 319-bp units, comprised of two related subunits. 122 bp of the first 172-bp subunit of the *Hinf*I repeat is 65% identical with a region of the salmon *Bgl*I repeat beginning at nt 237 (the start of the second subunit of Variant-2). Similarity between the two sequences was found to be statistically significant according to the computer program RDF2 (Pearson and Lipman 1988; Pearson 1990).

4.5 The repeat as a polymorphic population marker

Population specific markers have not been found for Atlantic salmon. Atlantic salmon show unusually low levels of genetic variation when assessed by techniques such as protein electrophoresis and RFLP analysis or direct nucleotide sequencing of a portion of the mitochondrial DNA cytochrome B gene (Davidson et al. 1989; McVeigh et al. 1991). On the other hand, simple sequence minisatellite probes show high levels of heterozygosity between salmon individuals (Taggart and Ferguson 1990a,b and Section 7.6).

Fig. 4.10. Sequence similarity of the Atlantic salmon *Bgl*I repeat (nt 216 to nt 399 of Variant-1, Fig. 4.5) with the Arctic charr *Dra*I repeat (ChDr7; Hartley and Davidson 1994; GenBank Acc. #L00991). Identical nt are defined by "|", missing nt by "-". Identity is 64% in a 183 nt overlap.

Single-digest restriction mapping of *Bgl*II repeat clones isolated from the Atlantic salmon genomic library detects intra-individual variation in the organization of repeats at different loci (Fig. 4.11). Variation in the length of individual repeats suggested that the *Bgl*II element might be used to detect restriction fragment length polymorphisms when *Bgl*II was used as the DNA hydrolyzing enzyme. Certain of these RFLPs might be fixed within different populations. To test this possibility, preliminary studies were carried out using a Variant-4 subclone as a probe in Southern blot analyses of DNA from Atlantic salmon of Scotland, Ireland and Newfoundland. Several bands show "allelic" variation between individuals within the same population (Fig. 4.12). Discrete population markers have not yet been identified, although the number of salmon tested was small. The *Bgl*II repeat could be useful for assessing the extent of genetic variation between different stocks.

4.6 Summary

(1) Sequence data from one tandem array of *Bgl*II repeats showed three variants of 403, 897, and 923 bp in length, and one variant truncated at its 3' end. The 403 bp unit is comprised of two related dimerized subunits.

(2) In Atlantic salmon, some repeats are found near rDNA genes, probably upstream from the 18S subunit in the IGS region.

(3) The *Bgl*II repeat is also strongly represented in the genome of brown trout, the only other *Salmo* species, but was not detected by Southern blotting DNA of other salmonids.

	10	20	30	40	50	60
ASBg122.9	CAAAC--TCCATAAATCAGCCAGGCCGGCCCCCATTTGACTTGGGCCCGTAGTAGGGTGCTGTCA					
CHDR7	TAAATTGGTCCAGACATC--TCAGGCCGGCCCCCTACTGACTTGTGCCCGCAGTATGTAGCTCCCA					
ASBg122.9	GGGTGGGTATGGAGGTCCCTATACCCCTTTAAAAGTATTTAAACAGTTATAAAATTTACCCCTG					
CHDR7	TAGCGGGCATTGAGGTGTGTAGCC-----ATTTAAACCA-TTATAAAATTGTGGGCCTG					
ASBg122.9	GTAACCCCAAATGC--TACGGACACGTCCCCACCAAAACATAGAATATTTCTGAGA					
CHDR7	GTAACCTAAAAAAATACC-AGGCGTACGGAAAAAACACCAGGTGCACGGGCTGG					

Fig. 4.11. Comparison of *Bgl*I repeat regions isolated from the Atlantic salmon genomic library. The segment that contained the *Bgl*I repeat element of each recombinant phage clone was subcloned into pTZ18R. Shown are *Bgl*I complete digests of DNA inserts from selected plasmid subclones (ASBgl series) separated on a 1.0% agarose gel. (All clones were first identified as containing portions of rDNA genes by PCR analysis, as shown in Fig. 4.7 and described in Section 4.3). Lane 1 contains size markers, and lane 2 a control PCR reaction containing no DNA. The subclones differ in the organization of their *Bgl*I repeat units. Length variants of the 923-bp band are seen. Also, the lowest band varies in intensity between samples, implying a difference in copy number.

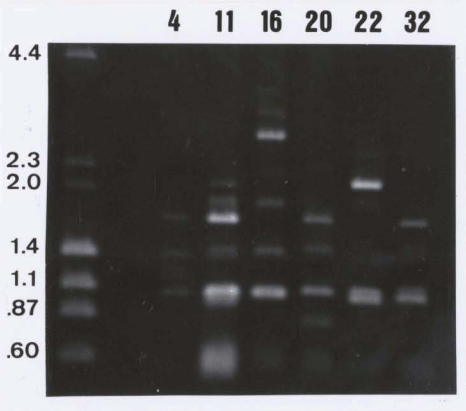
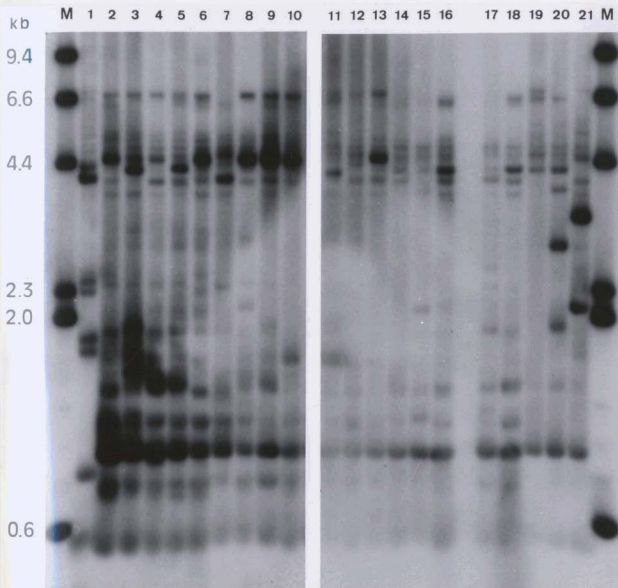


Fig. 4.12. Analysis of genomic DNA from a brown trout (lane 1) and Atlantic salmon from North America (lanes 2 to 13), Scotland (lanes 14 to 17) and Ireland (lanes 18 to 21). Approximately 3 µg of DNA were digested with *Bgl*II, electrophoresed in a 0.7% agarose gel, Southern blotted and probed with the insert of a Variant-4 subclone. Lane Brown trout; lane 1. Atlantic salmon: lanes 2 to 5, Bay D'Espoir, Newfoundland (NF); lane 6, LaHave R., Nova Scotia; lane 7, nonanadromous salmon, Gambo Pond, NF; lanes 8, 9, anadromous salmon, Gambo R., NF; lanes 10, 11, Black Brook, NF; lanes 12, 13, Grand Codroy R., NF. Lanes 14, 15, Almond R., Scotland; lane 16, Scourie Hatchery, Scotland; lane 17, Kames Fish Farm, Scotland; lanes 18, L. Melvin, Ireland; lanes 19 to 21, Swilly R., Donegal, Ireland. The markers are λ DNA cut with *Hind*III and labelled with [α -³²P]dCTP. Hybridization was under high stringency conditions, using the insert of subclone mpAlu18 as probe.



(4) The *Bgl*II repeat, when used as a probe in Southern blot analyses, detects restriction fragment length polymorphisms between individual fish.

CHAPTER 5. THE SALT1 REPEAT, A Tc1 TRANSPOSON-LIKE ELEMENT - RESULTS AND DISCUSSION

5.1 Isolation of a repetitive element identified by *Pst*I

When Atlantic salmon DNA is digested to completion with *Pst*I and fractionated on an agarose gel, two bands of approximately 580 bp and 800 bp are faintly visible upon staining with ethidium bromide (Fig. 4.1). In order to determine the nature of the corresponding repetitive element, DNA from the 580-bp band was isolated, ligated into pUC18 and the recombinant clones were screened by using the 580-bp band DNA as probe. When sequenced, 5 out of 16 hybridization positive, Atlantic salmon recombinant clones were found to contain similar sequences (Fig. 5.1). These are over 85% percent identical, although one clone, SSa100, has undergone a long 5' deletion.

Partial *Pst*I restriction digests of salmon genomic DNA and Southern analysis using the insert from clone SSa128 as probe failed to generate a ladder pattern of hybridization fragments, evidence that these repeats are not tandemly linked (data not shown). Rather, these segments probably derive from dispersed repeats possessing two internal *Pst*I sites.

The repeat sequences were compared against the EMBL DNA database using the FASTA sequence search server. Significant sequence similarity was found to a region encoding the carboxy-terminal of the putative transposases of the nematode transposable elements Tc1 of *Caenorhabditis elegans* (Rosenzweig et al. 1983) and Tcb1 (Barney) and Tcb2 of *C. briggsae* (Harris et al. 1990; Prasad et al. 1991). Similarity was also detected with *Drosophila* Tc1-like elements, Uhu (Brezinsky et al.

Fig. 5.1. Sequences of *Pst*I fragments of SALT1 repeat elements from various Salmonidae species: SSal, Atlantic salmon; OKis, Coho salmon (*Oncorhynchus kisutch*); OMyk, rainbow trout (*O. mykiss*); OMas, masu salmon (*O. masou*); SAlp, Arctic charr (*Salvelinus alpinus*); CLav, whitefish (*Coregonus lavaretus*). Nucleotides identical with the consensus are defined by ".", deletions by "-".

[illegible]

✓

1990), Minos (Franz and Savakis 1991), and Bari-1 (Caizzi et al. 1993).

Until recently transposons flanked by short inverted repeats were unknown in vertebrates. However, elements homologous with Tc1-transposons have been identified within an intron of a vasotocin gene in the primitive Agnathan hagfish (*Eptatretus stouti*; Heierhorst et al. 1992) and within an intron separating the transmembrane domains of the IgM locus of a teleost, the channel catfish (*Ictalurus punctatus*; Wilson et al. 1990; Henikoff 1992). Recently, homologous elements in the genomes of Atlantic salmon, rainbow trout and zebrafish have been sequenced by Radice et al. (1994).

5.2 Characterization of a full-length salmon Tc1 transposon-like element

The sequence of a complete Tc1-like element (designated SALT1-SSal1, SALmonid Transposon of *S. Salar*) was obtained from one clonal isolate of an Atlantic salmon genomic library. This sequence, together with flanking sequence (not shown here), has been deposited in the GenBank data base as accession #L22865. The predicted transposase (Tc1A) coding region is compared with that of the *C. briggsae* element Tcb1 (Harris et al. 1990) and the hagfish Tc1-like element, Tes1 (Heierhorst et al. 1992) in Fig. 5.2. The ORF predicted from SALT1-SSal1 begins at nt 207 with a second potential inframe start codon at nt 261. As with Tes1, the putative SALT1-SSal1 ORF contains a number of stop codons and is, therefore, probably defective. Ignoring these, and by forcing reading frame shifts at positions 769 and 1331, a 385-residue protein homologous to and terminating at the same position as the *Caenorhabditis* Tc1-like transposases is predicted. The longer ORF for Tes1 may have

Fig. 5.2. Sequence of the full length SALT1 element (SSal1) isolated from the Atlantic salmon genomic library, and comparison of its deduced putative transposase coding region with those of homologous elements from *E. stoufi* (Tes1; Heierhorst *et al.* 1992), *C. briggsae* (Tcb1; Harris *et al.* 1990), and the 5' flanking region of the *S. salar* ependymin gene (Ssepd; Müller-Schmid *et al.* 1992). Regions of similar inverted nucleotide sequence, designated ITR-A and ITR-B in the text, are underlined. Frameshifts have been introduced to the SALT1 sequence at nucleotides 769 and 1331 to maximize alignment of the amino acid sequences. Inserted gaps are indicated as "-", stop codons with "**", and conserved amino acid residues are in bold type. The intron region identified for Tc1 by Vos *et al.* (1993) and proposed for Tcb1 by Prasad *et al.* (1991) has been translated and is shown in lower case letters. The primers (a, b, and c) used for sequencing and for inter-SALT1 PCR (see Figs. 5.7 and 5.8) are indicated by lines above their sequence. A potential polyadenylation signal (Poly A) is shown. The complete sequence of SALT1-SSal1 has been deposited in the GenBank data base (accession #L22865).

120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

resulted from a mutation of the opal stop codon seen in SALT1-SSal1. In Tc1 a potential TATA box (TATAAAT) sits 47 bp upstream of the start codon. The only consensus TATA motif in SALT1-SSal1 is at position 80 (AATAAAT). There is a polyadenylation signal (AATAAA) at position 1490.

The existence of two exons encoding the TcA protein has been demonstrated empirically for Tc1 (Vos et al. 1993) and inferred for Tcb1 and Tcb2 (Prasad et al. 1991). The spliced transcript of Tc1 codes for 343 amino acids. For comparison with homologous regions of the fish Tc1-like elements, the presumptive intron sequence of Tcb1 has been translated in Fig. 5.2 and its encoded 20 amino acid residues and 3 stop codons are shown in lower case letters. It is unclear if the fish elements contain introns.

SALT1-SSal1 is flanked by imperfect inverted terminal repeats (ITRs), each commencing with 5'-TA-3'. This dinucleotide sequence has been identified as a target site for Tc1-like elements (Rosenzweig et al. 1983; Moerman and Waterston 1989) and is duplicated at the time of transposon insertion. Interestingly, immediately flanking the ORF is a short stretch of inverted repeat sequence (designated ITR-B) which has partial homology with sequence (ITR-A) near the very ends of the SALT1 element (Fig. 5.3, and described in Goodier and Davidson 1994a).

Very recently Z. Izsvak, Z. Ivics, and P.B. Hackett (unpublished) have entered in the GenBank database (Acc. Nos. L33469 to L33473) sequences of five Tc1 transposon-like loci from zebrafish (*Danio rerio*). Their entries of June 16, 1994 note the presence of unusually long inverted terminal repeats (<200 bp) in these Tc1-like

Fig. 5.3. Comparison of homologous sequences within the 5' and 3' inverted repeat sequences of SALT1-SSal1, (SSal1-ITR-A, flanking the element, and SSal1-ITR-B, flanking its open reading frame), and analogous 5' sequences from the salmon ependymin gene element (Ssepd-ITR-A and Ssepd-ITR-B). Identical nucleotides are shown by "|", and inserted gaps by "-".

```

Ssepd-ITR-B 5'      CAG-TG-GGTCAGAAGTTTACATACAC--TAAATT-GA
Ssepd-ITR-A 5'      TACAGTTGAAGTCGGAAGTTTACATACAC-TTAGGTTGGA
SSall-ITR-A 3'      TACAGTGCAT-TCGGAAGTATTTCAG-CCCCTTA-CTTT
SSall-ITR-A 5'      TAC-GTGCAT-TCGGAAGTATTTCAGACCCCTTGCCCTT
SSall-ITR-B 5'      TTTACATAAGTATTTCAGACCCCTTTCG-CT
SSall-ITR-B 3'      TTTACATAAGCATTTCAGACCCCTTTTA-CT

```

elements, each flanked at its 3' and 5' ends by 12 bp of similar sequence (Fig. 5.4A). This observation prompted a closer examination of the inverted repeat sequences from Atlantic salmon. The SALT1-SSal1 element also appears to contain long inverted repeats, but these are degenerate in their central portions. Thus, ITR-A and ITR-B are not separate repeat regions, as first inferred in Goodier and Davidson (1994a), but the termini of a single pair of long imperfect inverted repeats extending from the ends of the element as far as the predicted amino acid coding region.

A search of the EMBL data library has revealed a portion of the 5' flanking region of an ependymin gene of Atlantic salmon sharing homology with the SALT1 sequence. The published sequence (Müller-Schmid et al. 1992) includes the first 972 bases of a SALT1-like element inserted in an inverse orientation 295 bp from the anticipated transcription start point of the ependymin gene (Fig. 5.2). The SALT1-like element commences with 36 nucleotides having similarity with the ITR-A sequence flanking SALT-SSal1; these are imperfectly repeated about 100 bases downstream (the 5' ITR-B; Fig. 5.3). The absence of known transcription elements upstream of the the TATA box in this *S. salar* ependymin gene locus, together with its "very specific spatial and temporal expression" (Müller-Schmid et al. 1992), prompt speculation that expression of at least one salmon ependymin gene may be influenced by an inserted transposable element.

The ependymin gene element shows considerable similarity with salmonid Tc1-like elements isolated by Radice et al. (1994): over 81% amino acid sequence identity with three loci from Atlantic salmon (GenBank accession nos. L12206 to L12208) and

Fig. 5.4. Alignment of the 5' and 3' flanking sequences of A) ZeffTrana, *Tel*-transposon-like sequence from *Danio rerio* (Z. Izsvak, Z. Ivics, and P.B. Hackett, unpublished, GenBank Acc. #L33470), B) SALT1-SSall, and C) *SmoTel*lt (Radice et al. (1994); #L12206).

ZefTrana 5' CAGTGAAGTCGGAAGTTACATACGCTGTATAAAAAGGCACATAACCATTACTAAAAAA
ZefTrana 3' CAGTTGAAGTCGGAAGTTACATACACACTAAAAAAGGAATATAACCATTTTTTGTTAA

ZefTrana 5' A-GTCAGATGTTAA-TGTGACTAAACTTTTCTCTTTTAGSTAAGTCAGGATTATC--AC
ZefTrana 3' ATGCTCGATGGTAAATCATACTAAACGTTTACTATTTTAGSTCCGTTAGGATTACCTAAA

ZefTrana 5' GTGTTTATGTTCTGCTTAATAGCAGAATAATGAGAGAGAGATTTTTCAGAAATCGTTAT
ZefTrana 3' CCATTTACATA-TGCTAAATGCCAGAATAATGAGAGA-AG-TTTTAAAAATAAATTTCT

ZefTrana 5' AA-CTTTTCTTGAAAGTC--AAGTTTACACA
ZefTrana 3' ATTAATTTCTAGACTTTTAAAGTTTACATA

[illegible][illegible]

78% identity with one locus of rainbow trout (*Oncorhynchus mykiss*; accession no. L12209). SALT1-SSal1, on the other hand, is less than 32% identical with the amino acid sequences predicted for these four other salmonid elements. These complete salmonid elements, like SALT1-SSal1, are flanked by long degenerate inverted repeats which contain homologous ITR-A and ITR-B regions, and these are very similar to the corresponding 5' repeats of the ependymin gene element. The ITR sequences from one element (acc. #L12206) of Radice *et al.* (1994) is shown in Fig. 5.4C

The CAGTGY consensus sequence, noted by Henikoff (1992; see also Collins and Anderson 1994) as present in the inverted terminal repeats of members of the Tc1-family, is also found in the SALT1-SSal1 ITRs. However, this motif varies both in sequence and position between different fish species. Those elements identified by Radice *et al.* (1994) carry a variant CAGTTGA sequence, while the motif may be missing entirely from the hagfish inverted repeats (the sequence CAGTTT is present). The 85-bp ITRs of the catfish element, identified by Henikoff (1992) as a Tc1-like homologue, are the most divergent. The CAGTGY motif is present, but internal to the ITR rather than at its extreme 5' end. The catfish transposon has, in fact, inserted into an A-T rich region making identification of its termini uncertain.

Comparison of homologous sequences within the inverted repeat regions of known fish Tc1-like elements identifies a consensus sequence (Fig. 5.5). Nucleotides at certain positions are highly conserved. Sequence in the ITRs of some transposable elements have been shown to be required *in cis* for excision and transposition (Karess and Rubin 1984). A domain mapped within first 63 residues of the amino-terminal end

Fig. 5.5. Comparison of homologous sequences of the inverted 5' and 3' repeat sequences flanking fish Tc1-like elements in the channel catfish IgM locus (IpTc1-ITR-A; Wilson et al. 1990) and the hagfish vasotocin gene (Tc1-ITR-A; Heierhorst et al. 1992), SALT1-SSal1 (SSal1-ITR-A) and its open reading frame (SSal1-ITR-B), the salmon cpendymin gene element (Ssepd-ITR-A and Ssepd-ITR-B; Müller-Schmid *et al.* 1992), and elements isolated by Radice et al. (1994) from several fish species: zebrafish (ZefTc1ltd), Atlantic salmon (SmoTc1lt, #L12206; SmoTc1lta, #L12207; and SmoTc1ltb, #L12208), and rainbow trout (SmoTc1ltd, #L12209). Inserted gaps are indicated as "-".

CCSC2505

Ipfel-ITR 5'

Ipfel-ITR 3'

Tes1-ITR 5'

Tes1-ITR 3'

Zetf1c1b-ITR 5'

Zetf1c1b-ITR 3'

Ssall-ITR-A 5'

Ssall-ITR-A 3'

Ssagp-ITR-A 5'

Ssagp-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

Ssotc1b-ITR-A 5'

Ssotc1b-ITR-A 3'

TACGTTGA-GTCGAGTTTACATCA-CTAG-TT

TTTGATTTTGTGTTTACGATTAAGACAGCAACATTGATCATCTTGGAGATGCTGCTATATAGTT-GBATAGT

TTTATTTATTTATTTTGGACATTTGGAGACAGCAACATTGATCATCTTGGAGATGCTGCTATATATAGTT-GBATAGT

TACTCTACGCTGGAAGTTTGAAGACGCCCATTTTCAGTTTCTACTGAAATTAAGCGTT

TACTCTACGCTGGAAGTTTGAAGACGCCCATTTTCAGTTTCTACTGAAATTAAGCGTT

TACGTTGA-ACGAGAGTTTACGAGCTGTAAAGAGAGCTAGATTTTAA

TACGTTGA-ACGAGAGTTTACGAGCTGTAAAGAGAGCTAGATTTTAA

TAC-GTTCAT-ACATAGCTTTCTATGCTCTGAAAG-ACATACCATTTAA

TAC-GTTCAT-ACATAGCTTTCTATGCTCTGAAAG-ACATACCATTTAA

TACGTGCAT-TGGAGTAT-AGGCC-CTA-CTT

TACGTGCAT-TGGAGTAT-AGGCC-CTA-CTT

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

TACGTTGA-GTCGAGTTTACATCA-CTAGTTGA

of nematode Tc1 transposase specifically binds to ITR sequence (Vos et al. 1993).

5.3 Species distribution of SALT1 elements

To determine the phylogenetic distribution of the SALT1 Tc1-like element, genomic DNAs from salmonid and other fish species, frogs (*Anura*), bird, and mammal species were digested to completion with *Pst*I and subjected to Southern blotting. These were probed with clones SSa19, SSa29, or SSa44 to examine the phylogenetic distribution of the *Pst*I repeat (Fig. 5.6). Blots hybridized and washed at high stringency showed strong hybridization to all salmonids but to no other species; the pattern of hybridization was similar in each case. The ethidium bromide staining 580-bp *Pst*I bands from selected *Salmo*, *Salvelinus*, and *Oncorhynchus* species were isolated and sequenced as described above for Atlantic salmon (Fig. 5.1). Surprisingly, intra-individual sequence variation of element copies was no less than inter-species variation. Based on phylogenetic analysis of the sequence data set using the sequence analysis program PHYLIP, no species specific sequence markers could be detected nor were any subgroupings of sequences obvious (Fig. 5.7). Some of the sequenced fragments shown in Fig. 5.1 are interrupted by long internal deletions. (In the case of *Drosophila* P elements, Engels et al. (1989) have suggested that internal deletions arise from inefficient repair of double-strand breaks occurring at the time of excision). None contain a complete ORF. They derive from elements probably incapable of autonomous transposition and therefore it is unlikely that they have been under selective constraint to retain sequence homogeneity. In the case of the SALT1 elements, rates of 'concerted

Fig. 5.6. Zoo blot analysis of the distribution of SALT1-like elements in fish and anuran species. Clone SALT1-SSal44 was the probe. Lanes 1, Atlantic salmon; 2, masu salmon; 3, rainbow trout; 4, whitefish. Lanes 1 to 4, autoradiography was for 6 h. Lanes 5, sea lamprey; 6, dogfish shark; 7, skate; 8, Atlantic cod; 9, winter flounder; 10, capelin; 11, cunner; 12, *Xenopus laevis*; 13, *Rana* sp. Lanes 5 to 13, autoradiography was for 7 d. Hybridization was at low stringency (see Section 3.4). Other salmonid species showed hybridization patterns similar to lanes 1 to 4, and included brown trout, Arctic charr, chinook salmon, pink salmon, coho salmon, chum salmon, Japanese charr, brook trout, and Arctic grayling. Other animal species tested and showing no DNA hybridization included gannet, quail, chicken, rabbit, moose, caribou, black bear, mouse, baboon, and human.

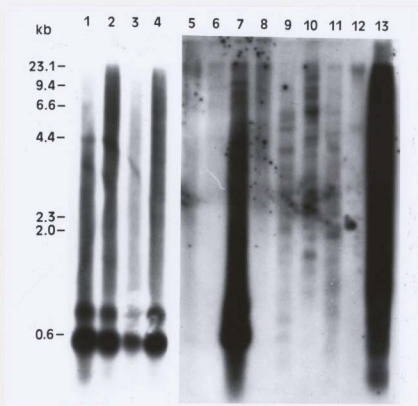
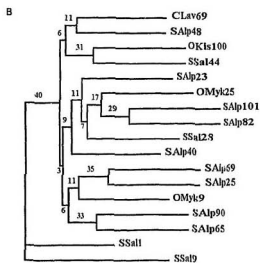
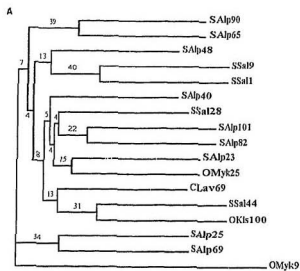


Fig. 5.7. Relationships of the SALT1-like sequences isolated from different Salmonidae species analyzed with the phylogenetic package PHYLIP (see Fig. 5.1 for symbols). Sequences with deletions of over 65 bp were omitted from the analysis (SSal100, SSal40, OKis83, OMas46, and Clav5). The tree in A) was generated with a nearest neighbour-joining algorithm (program NEIGHBOR), and in B) with Fitch-Margoliash and least squares methods (program FITCH). Numbers above the branches refer to the number of times each branch was supported after 100 bootstraps (program SEQBOOT).



evolution' (Arnheim 1983), processes which encourage homogenization of the members of a repeat family within a single species, appear to have been low. The sampling reported here was limited, however, and more extensive sequence data are required if the existence of different element lineages or of phylogenetic relationships is to be assessed (see Batzer et al. (1993) and VanderWiel et al. (1993) for discussion of this point).

Under low stringency conditions, hybridization with DNA of Atlantic sturgeon (*Acipenser oxyrinchus*; not shown in Fig. 5.6), skate (*Raja* sp.), and *Rana* sp. was detected after 6 hr autoradiographic exposure. After exposure for 7 days a much fainter signal also was visible as a series of bands within the lanes for winter flounder (*Pleuronectes americanus*), capelin (*Mallotus villosus*) and cunner (*Tautoglabrus adspersus*). In contrast to the strong signal of *Rana*, probe hybridization to *Pst*I-digested *Xenopus laevis* DNA occurred primarily as two faint bands, approximately 4.8 kb in length, that were also visualized by ethidium bromide staining. Hybridization with DNA of the Agnathan sea lamprey (*Petromyzon marinus*), the Chondrichthyan dogfish (*Squalus* sp.) and Atlantic cod appeared to be non-specific.

In Arctic charr, immediately 5' of a locus containing a *Mbo*I repeat family element (ChMb12; GenBank Acc. #L01073) characterized by Hartley and Davidson (1994), there is sequence unrelated to the *Mbo*I repeat and having 89% identity to a 61-bp stretch of SALT1-SSa11 (commencing at nt 1200). Members of the *Mbo*I family are tandemly repeated and consist of a 33-mer doublet (Fig. 5.8). Interestingly, this repeat has been found to be missing from an isolated population of *S. alpinus* from Wester

Fig. 5.8. Sequence similarity of SALT1-SSal1 (nt 1201 to nt 1262, Fig. 5.2) with the Arctic charr *Mbol* repeat (ChMb12; Hartley and Davidson 1994; GenBank Acc. #L01073). Identity is 92% in a 61 nt overlap. Identical nt are defined by "|", missing nt by "-". Also shown is the alignment pattern of the two *Mbol* repeats present in ChMb12.

-----|-----|-----|-----|-----|-----
 10 20 30 40 50 60

SALT1-SSalI
CATCTCTGGAGAGACTTGAAAATATCTGTGCAGCAACGTTCCCATCCAACC

ChMb12
GATCCGAACATGTGTGGAAGAGAGAGACCTGAAAATAGTGTGCAGCGACG--CTCCATCCAACC
*Mbo*I*Alu*I

SALT1-SSalI
TGACAGAGCCTGAGAGGATCTGCGTAGAA

ChMb12
TGACAGAGCCTGAGAAGACTTACTCTCTCTATTATATAATGACAGACCAGCTAG-TC >
*Alu*I

ChMb12
*Mbo*I

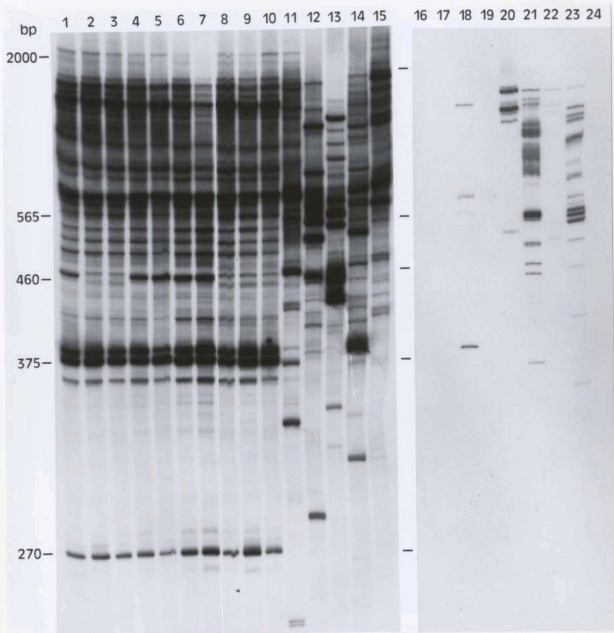
Ross, Scotland, but present in other Scottish populations (noted in Hartley and Davidson (1994) as unpublished data). One may conjecture that the precursor *Mbo*I repeat sequence, prior to its amplification in Arctic charr, was lost from the ancestral genome of the Wester Ross population, perhaps as a consequence of excision during a transposition event involving a SALT1 element.

5.4 The genomic distribution of SALT1 elements

The technique of human *Alu* PCR utilizes a single primer directed outwards from the *Alu* repeat sequence to amplify DNA between elements inversely oriented with respect to one another (Nelson et al. 1989). In an analogous fashion, SALT1 primer a or c (Fig. 5.1) was end-labeled for use individually in PCR-based amplifications of genomic DNA from Atlantic salmon of both North American and European populations, and from other salmonid species. Amplified products were resolved on a denaturing polyacrylamide sequencing gel. Following autoradiography, a large number of products were detected, generating patterns similar to those of DNA fingerprinting gels, and indicating close proximity of SALT1 elements at some loci (Fig. 5.9). Bari-1, a recently characterized Tc1-like element from *D. melanogaster*, has been found to be largely clustered in regions of heterochromatin (Caizzi et al. 1993).

Interspecific variations in inter-SALT1 PCR banding patterns were numerous with primer a, but only a single fragment length polymorphism was detected among 16 Atlantic salmon individuals chosen from Canada, Ireland and Scotland (Fig. 5.9). Similar results were obtained with primer c, although single band shifts were detected

Fig. 5.9. Autoradiograph of inter-SALT1 PCR of genomic DNA from Atlantic salmon (lanes 1 to 10), other salmonids (lanes 11 to 15), and non-salmonid species (lanes 16 to 23). ^{32}P -labelled primer a was used alone in the PCR reactions. One sample, containing degraded DNA was not included in this figure (between lanes 7 and 8). Amplification products in lanes 1 to 10, and lanes 11 to 23 were resolved on separate gels. Atlantic salmon: lane 1, Bay d'Espoir, Newfoundland; lanes 2, 3, LaHave R., Nova Scotia; lane 4, St. John's R., New Brunswick; lane 5, Black Brook, Newfoundland; lanes 6, 7, Ungava Bay, Quebec; lanes 8, 9, Almond R., Scotland; lane 10, Swilly R., Donegal, Ireland. Brown trout, lane 11; chum salmon, lane 12; Japanese charr, lane 13; masu salmon, lane 14; rainbow trout, lane 15; sea lamprey, lane 16; dogfish shark, lane 17; skate, lane 18; sturgeon, lane 19; winter flounder, lane 20; *Rana* sp, lane 21; mouse, lane 22; human, lane 23; control sample containing no DNA, lane 24. A fragment length polymorphism of approximately 460 bp is seen in Atlantic salmon. Fragment sizes were determined by comparison with a sequencing reaction electrophoresed at the same time.

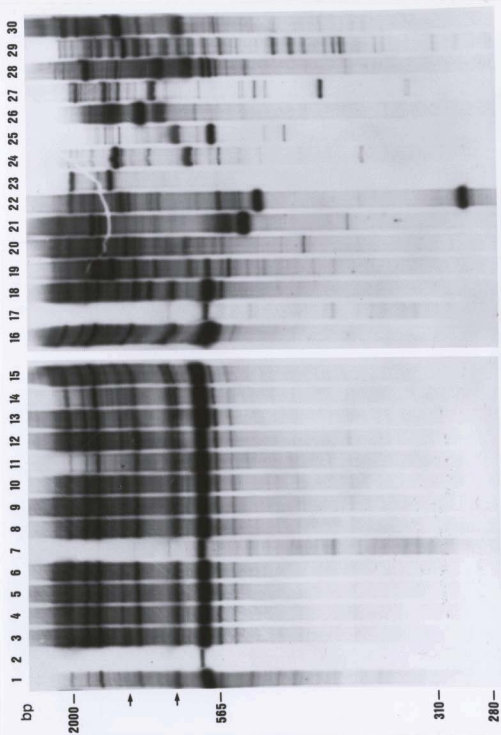


in several individuals (Fig. 5.10). These results are consistent with the hypothesis that extant Atlantic salmon populations are recently derived from a common population and have colonized their present habitats since the last ice-age (see Davidson et al. 1989). These results also suggest that either insufficient time has elapsed for new transpositions to have become fixed in the different populations or else these transposons are rarely active.

Using primer *a*, inter-SALT1 PCR amplification products were obtained from non-salmonid species including some other fish, an amphibian, mouse, and human. No products were detected for dogfish shark, while those of lamprey and sturgeon were faint and are not discernable in Fig. 5.9. However, primer *c* generated products ranging in number from 3 (lamprey) to over 20 (mouse), for every non-salmonid species examined (Fig. 5.10). Some of these results are at variance with Southern blot analysis. Although one cannot rule out the possibility that these primers may be generating RAPDs (see Section 1.5.1) in some instances, the length of the primer and the high annealing temperature of the PCR suggest that SALT1-related elements are being detected. Tc1-like elements appear, therefore, to be widespread in vertebrate genomes.

SALT1 elements have achieved high copy number in salmonids. As shown by densitometric analyses of ethidium bromide-stained gels, the two bands visible when Atlantic salmon genomic DNA is digested with *Pst*I (Fig. 4.1) account for over 0.2% of the genome. Furthermore, when 3,600 recombinant clones of the λ EMBL3 genomic library were screened using *Ssa*28 as probe, 335 positive plaques were detected following 4 days of autoradiographic exposure. If the average length of a phage insert

Fig. 5.10. Autoradiograph of inter-SALT1 PCR of genomic DNA from Atlantic salmon (lanes 1 to 17), other salmonids (lanes 18 to 22), and non-salmonid species (lanes 23 to 30). ^{32}P -labelled primer c was used alone in the PCR reactions. Atlantic salmon: lanes 1, 2, Bay d'Espoir, Newfoundland; lanes 3, 4, LaHave R., Nova Scotia; lanes 5 to 7; Gambo, Newfoundland; lanes 8, 9, Black Brook, Newfoundland; lane 10, Grand Codroy, Newfoundland; lane 11, Scotland; lane 12, Lussa R., Scotland; lane 13, Scourie fish hatchery, Scotland; lane 14, Scotland; lane 15, Owentogher R., Ireland; lane 16, Swilly R., Donegal, Ireland; lane 17, L. Melvin, Ireland. Brown trout, lane 18; masu salmon, lane 19; Arctic charr, lane 20; grayling, lane 21; whitefish, lane 22; sea lamprey, lane 23; dogfish shark, lane 24; skate, lane 25; sturgeon, lane 26; winter flounder, lane 27; *Rana* sp, lane 28; mouse, lane 29; human, lane 30. Lanes 1-15 and lanes 16-30 were run on separate gels under the same conditions. A control sample containing no DNA was also included and showed no signal (not shown in this figure). Fragment length polymorphisms are indicated with arrows. The highly variant fragment lengths in lanes 2, 7, and 17 may be artifacts caused by insufficient DNA being included in the PCR reaction. Fragment sizes were determined by end-labelling *Hind*III digested λ DNA and ϕX174 DNA cut with *Hae*III.



is taken as 15 kb, and one assumes that there is only one SALT1 homologue per positive clone, then SALT1 elements comprise 0.9% of the Atlantic salmon DNA in this library. These results translate into approximately 15,000 copies of the Tc1-like element per haploid genome (based on a "c" value of 3.0 pg for salmonid fish; Hinegardner and Rosen 1972). The high copy number in salmonids could have arisen through chromosome duplication (Ohno et al. 1969; Ohno 1970), as well as transposition.

The inferred amino acid sequence of the transposase of SALT1-SSal1 is more similar to that of nematode Tcb1 than to the corresponding sequence of hagfish Tes1 (32% and 26% identity, respectively). This observation, together with the uneven occurrence of homologues as determined by Southern blot hybridization might suggest that horizontal transmission to some fish species has occurred. On the other hand, Tc1-like transposons may be pervasive in vertebrates, and only elements which have become differentially amplified in some lineages have been detected. Interpretation of the evolution of this group of transposons is further complicated by the possible existence of several subfamilies, as seems to be the case in fish.

5.5 Summary

(1) A full length interspersed element (designated SALT1-SSal1), having internal homology to the transposase (Tc1A) coding region of the Tc1 family of transposons, was isolated and characterized. The element is 1535 bp long and flanked by degenerate inverted repeats whose 5' and 3' ends are similar in sequence to each

other.

(2) Homologous *PstI* fragments within the putative transposase coding regions were isolated from an Atlantic salmon and individuals of other salmonid species. There was considerable intraindividual as well as interspecies sequence variation in these Tc1 transposon-like elements.

(3) The full-length *S. salar* element described here is surprisingly divergent from three other salmon Tc1-like sequences independently isolated by Radice et al. (1994).

(4) Preliminary data suggest that this family of repeats is distributed among many members of several diverse vertebrate lineages. Selective amplification or loss of elements in certain lineages appears to have occurred.

CHAPTER 6 THE *NheI* REPEAT - RESULTS AND DISCUSSION

6.1 Sequences of Atlantic salmon *NheI* repeats (ASNHEs)

This chapter presents the characterization of a 380-bp repetitive element that is flanked by the recognition sequence for the restriction enzyme *NheI* and is prevalent within the genomes of the two species in the genus *Salmo*, Atlantic salmon and brown trout. Members of this repeat family were cloned into pUC18, and the sequences of 16 representative Atlantic salmon loci (ASNHEs) are shown in Fig. 6.1. The consensus sequence (entered in GenBank as Acc. #L25409) is 378 bp in length, but individual variants range from 367 bp (ASNHE8) to 388 bp (ASNHE23). The *NheI* repeats contain three perfect (1D, 3D, 4D) and two imperfect (2D, 5D) direct subrepeats, and two perfect (1I, 4I) and four imperfect (2I, 3I, 5I, 6I) inverted subrepeats. The latter include palindromes and quasipalindromes (repeats separated by a short spacer) which could potentially form cruciform structures. Short internal repeats have often been reported in satellite DNA (Miklos and Gill 1982) and in SINE-like repeats (Kido et al. 1991). Several major deletion or insertion events have occurred and all are located at internal subrepeat sites. The role of repeats in promoting local sequence excision has been investigated in detail in prokaryotes (Collins 1980; Trinh and Sinden 1993) and postulated to be a source of genetic instability in eukaryotes, possibly due to slipped-strand mispairing between stretches of homologous sequence (Gordenin et al. 1993; Ruskin and Fink 1993, and references therein).

The ASNHEs are A+T-rich (64%) and this is similar to satellite DNA from carp (63%; Datta et al. 1988) and pollock (65%; Denovan and Wright 1990). Individual

Fig. 6.1. Comparison of the nucleotide sequences of 16 cloned *NheI* Atlantic salmon repeats (ASNHE) and their corresponding consensus sequence (AS Cons) shown above. Nucleotides identical with the consensus sequence are shown by ".". Inserted gaps are indicated as "-". The symbols 1I to 6I and 1D to 5D marked above the ASNHE consensus sequence identify inverted and direct repeats, respectively. Recognition sites for restriction enzymes are indicated above the consensus sequence, and those that are present in only some variants are shown in brackets.

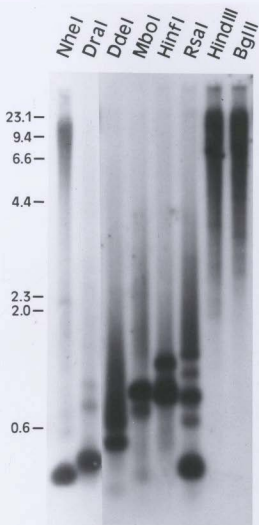
copies are very similar to one another and 95% to 98% (the average is 96.3%) identical to their consensus sequence (deletions or insertions being counted as single events). This too appears to be a characteristic of DNA repeats from other fish species: pollock (94% identity for 6 copies, Denovan and Wright 1990); zebrafish (95% identity for 13 copies, Ekker et al. 1992); and tilapia (98% identity for 7 copies, Wright 1989).

6.2 Genome organization of the Atlantic salmon *NheI* repeat

Densitometric analyses of photographic negatives of ethidium bromide stained agarose gels indicated that the DNA constituting the *NheI* 380-bp band peak accounted for an average of 1.2 (± 0.2)% of the total genomic DNA of three different salmon from western Newfoundland rivers. Assuming a "c" value of 3.0 pg (Hinegardner and Rosen 1972), this suggests that there are approximately 10^5 copies of this repeat in the haploid genome of Atlantic salmon.

Single sites for the restriction endonucleases *DraI* and *HaeIII* and two sites for *HpaII*, while present within the consensus sequence, have been lost from some copies due to point mutations. *NheI* and *DraI*, together with *DdeI* and *MboI* (found in all 16 copies), *RsaI* (found in only a small number of copies), and *HinfI*, *HindIII* and *BglII* (absent from all 16 copies sequenced) were used in a Southern blot analysis of genomic DNA from a single Atlantic salmon. In the case of maxi-satellite DNA and tandemly arranged repeats, one would expect to see a long ladder pattern of fragments due to random loss of restriction sites from some monomer units. In Fig. 6.2, such a pattern was clearly evident only for the *RsaI* digest, although the bands were not multimers of

Fig. 6.2. Genomic organization of the *NheI* repeat. Atlantic salmon DNA from a single individual was digested with the restriction enzymes shown and subjected to Southern blot analysis using the insert of plasmid clone ASNHE6 as the probe. The size markers in kb are derived from the migration of λ DNA cut with *HindIII*.



380 bp. Southern blot analysis of partial *NheI* digests of salmon DNA similarly failed to generate hybridization patterns typical of tandemly repeated "satellite-like" DNA (Fig. 6.3). Following long exposure of the Southern blots depicted in Figs. 6.5 and 6.9 (see below), faint hybridization bands above the *NheI* monomer can be seen against an intense background hybridization smear. Nevertheless, it appears that *NheI* repeats are either arranged in very short tandem arrays or else represent internal fragments of a longer, as yet unidentified, interspersed repeat.

6.3 Intra-specific variation of the *NheI* repeat

As repeats have proven useful for detecting polymorphisms in fish DNA (see Section 1.5.2), I tested for intraspecific genetic variation detectable by the *NheI* element. Genomic DNA from a total of 22 Atlantic salmon originating from the LaHave R. in Nova Scotia, six river systems of Newfoundland, four river systems of Scotland, and five of Ireland was digested with the restriction enzymes *HindIII*, *HinfI*, *MboI* or *RsaI* and subjected to Southern blot analysis. No variation in banding patterns of the *NheI* element was observed in these individuals from widely distributed populations (Fig. 6.4). On the other hand these individuals show considerable RFLP variation when analyzed with probes from the *BglI* multi-locus repetitive DNA (see Section 4.5) and from two minisatellite loci of Atlantic salmon (Sections 7.5 and 7.6). The *NheI* repeat, therefore, does not appear to be useful for population genetic studies in Atlantic salmon.

Fig. 6.3. Southern blot analysis of partial digests of salmon genomic DNA probed with the DNA insert from ASNHE6. Aliquots in lanes 2 to 10 were digested for 1 hr at 37°C, with varying concentrations of *NheI*, while digestion for lane 1 DNA was permitted to continue overnight. Samples were electrophoresed in a 1.0% agarose gel, and blotted to Biodyne nylon membrane. Hybridization was performed under high stringency conditions using the modified DNA hybridization protocol of the RCMP, Ottawa (conditions were as described in Section 3.5, with the exception that washing was at 65°C). Lanes from left to right: 20, 10, 5, 2.5, 1.25, 0.62, 0.31, 0.16, 0.08, and 0.04 units of *NheI*.

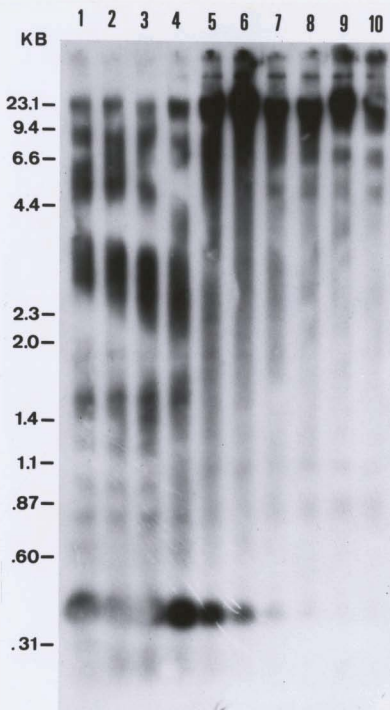
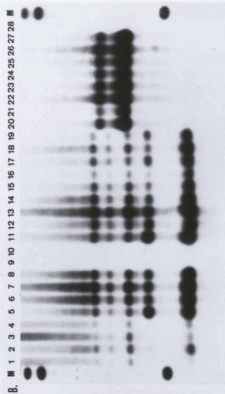
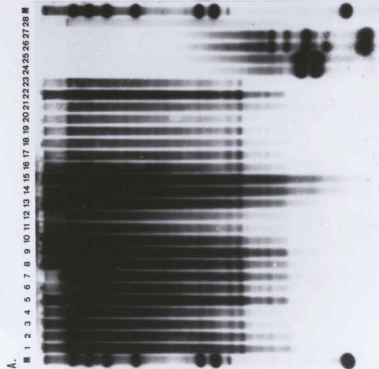


Fig. 6.4. Southern blot analysis of Atlantic salmon DNA from Newfoundland (NF), Nova Scotia (NS), Scotland (SCOT), and Ireland (IRE) digested with *HindIII*, *Sau3A*, *RsaI* and *HinfI* and probed with the insert of brown trout clone BTNHE68: A) lanes 1 to 23, *HindIII*; lanes 24, 25, *Sau3A*; lanes 26, 27, *RsaI*; lane 28, degraded DNA; B) lanes 1 to 19, *RsaI*; lanes 20 to 28, *HinfI*; C) lanes 1 to 10, *HinfI*; lanes 11 to 28, *Sau3A*. Hybridization was performed under high stringency conditions using the modified DNA hybridization protocol of the RCMP, Ottawa (conditions were as described in Section 3.5, with the exception that washing was at 55°C).

A) Lanes 1,2, Bay D'Espoir, NF; lanes 3,4, LaHave R., NS; lanes 5 to 7, Gambo, NF; lane 8, Black Br., NF; lane 9, Little R., NF; lane 10, Grand Codroy, NF; lanes 11, 12, East Placentia, NF; lanes 13, SCOT; lane 14, Lussa R., SCOT; lane 15, Scourie Fish Hatchery, SCOT; lane 16, SCOT; lane 17, Owentogher R., Donegal, IRE; lane 18, Keewagh R., Donegal, IRE; lanes 19, 20, Swilly R., Donegal, IRE; lane 21, L. Melvin, IRE; lane 22, Swilly R., IRE; lane 23, Glen R., Donegal, IRE ; lanes 24 to 27; East Placentia, NF. B) Lane 1, Bay D'Espoir, NF; lane 2, LaHave R., NS; lanes 3 to 5, Gambo, NF; lane 6, Black Br., NF; lane 7, Little R., NF; lane 8, Grand Codroy, NF; lanes 9, 10, East Placentia, NF (DNA was degraded); lane 11, SCOT; lane 12, Lussa R., SCOT; lane 13, Scourie Fish Hatchery, SCOT; lane 14, Owentogher R., IRE; lane 15, Keewagh R., IRE; lanes 16 to 17, Swilly R., IRE; lane 18, L. Melvin, IRE; lane 19, Glen R., IRE (DNA was degraded); lanes 20 to 28, same as lanes 1 to 9. C) lanes 1 to 10, same as A) lanes 10 to 19; lanes 11 to 28, same as A) lanes 1 to 19. Size markers are *HindIII*-cut λ DNA.



6.4 Species distribution of salmon *NheI*-like sequences

The distribution in other fish species of DNA sequences related to the Atlantic salmon *NheI* repeats was determined by Southern blot analysis. There was strong hybridization of the probe to DNA from brown trout. A weak signal was detected (after low stringency washes) with DNA from other Salmoninae species but this was more intense than the faint, and probably non-specific, background hybridization exhibited by DNA from the non-Salmoninae fishes, including the Salmonidae species grayling and whitefish (Fig. 6.5). Therefore, it appears that this repetitive element is found primarily in the genus *Salmo*.

6.5 Sequence analyses of *Salmo* *NheI* repeats

Seven hybrid recombinant plasmids containing *NheI* repeats from brown trout were isolated and the sequences of the inserts were determined (Fig. 6.6). The relationships among the sequences of the sixteen ASNHES and the seven brown trout *NheI* repeats (BTNHES) were examined using the PHYLIP package. Fig. 6.7 shows the results of this analysis and indicates that the individual repeats do not fall into species-specific groups. Rather, it appears that there are subfamilies of *NheI* repeats and these families are found in both species of the genus *Salmo*. This result, taken with the observation that the *NheI* repeat is not a major element in *Oncorhynchus* or *Salvelinus* species, suggests that there was an expansion of the *NheI* family of sequences in the ancestor of Atlantic salmon and brown trout after the split between *Salmo* and the other genera in the Salmonidae.

Fig. 6.5. Zoo blot analysis of the taxonomic distribution of *NheI* elements. Genomic DNA was digested with *NheI* and the insert from clone ASNHE6 was used as the probe. Lanes 1, Atlantic salmon; 2, brown trout; 3, coho salmon; 4, chum salmon; 5, masu salmon; 6, rainbow trout; 7, Arctic charr; 8, Japanese charr; 9, brook trout; 10, grayling; 11, whitefish; 12, sea lamprey; 13, dogfish shark; 14, Atlantic cod; 15, cunner; 16, *Rana* sp.; 17, gannet; 18, caribou; 19, human. The restriction enzyme has failed to hydrolyze brook trout DNA (lane 9). Hybridization was at low stringency (see Section 3.5).

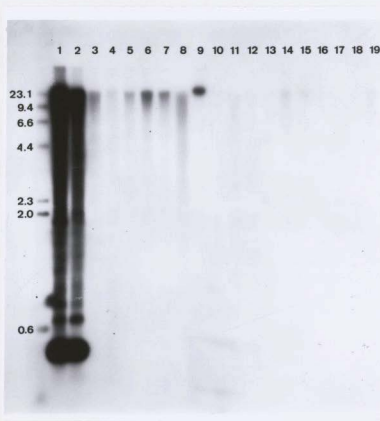
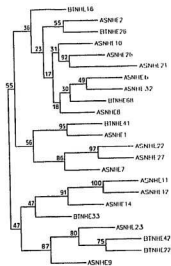


Fig. 6.6. Comparison of the nucleotide sequences of 7 cloned *NheI* Brown trout salmon repeats (BTNHE), their corresponding consensus sequence (BT Cons) shown above, and the ASNHE consensus (AS Cons; see Fig. 6.1). Dots indicate that the sequence is identical to that of the BTNHE consensus and dashes indicate deletions.

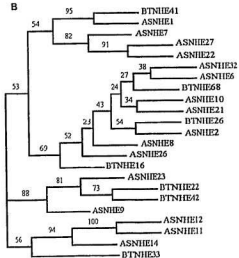
	10	20	30	40	50	60	70	80	90	100
AS Cons
BT Cons	G..TAGCATATAATGTCACATAAAC-CCAAACACAGCTAAATGCAG	CTACC-TTTGATGATCTTCATGATGACACTCCTAGACATATAGTTATA								
BTNRH16C.....T.....A.....
BTNRH26A.....A.....A.....
BTNRH58A.....T.....A.....A.....AA.....
BTNRH41T.....A.....
BTNRH33T.....C.....CA.....
BTNRH22A.....
BTNRH42A.....G..AG.....
AS ConsC.....T.....
BT Cons	CAATACATGCATGTTTTGTCAATCAAGTCATATTATATCAAAAAGCTTTT-ACATTAGCATGTCG	-----GTTCCAGAACTATGATACCCAC								
BTNRH16C.....
BTNRH26C.....
BTNRH58TAGC.....T.....
BTNRH41C.....TAGCAT.....G.....T.....
BTNRH33A.....GC.....T.....C.....
BTNRH22C.....
BTNRH42C.....
AS ConsA..C.....
BT Cons	C-GCAAACTTCCGGTGATTACTAAA-----TTACTCAGGATAAGCTTCACAAAAA-CATACATATATTTAAGAAATTATATACATAA									
BTNRH16
BTNRH26
BTNRH58
BTNRH41
BTNRH33A..C.....
BTNRH22
BTNRH42G.....C-ATTACTAAA.....G.....TT.....
AS Cons
BT Cons	CTCCTTATGCACTCGCTATGTCGATTTTAAATAGCTTTTCGGTAAAGCAGCATTTGCAATATCTGAGTAGATATCCGCCATCAGAGGCTACG									
BTNRH16A.....GC.....
BTNRH26
BTNRH58
BTNRH41
BTNRH33
BTNRH22
BTNRH42

Fig. 6.7. Relationships of the ASNHE and BTNHE sequences analyzed with the phylogenetic package PHYLIP. The tree in A) was generated with a nearest neighbour-joining algorithm (program NEIGHBOR), and in B) with Fitch-Margoliash and least squares methods (program FITCH). Numbers above the branches refer to the number of times each branch was supported after 100 bootstraps (program SEQBOOT).

A



B



A comparison of the ASNHE consensus sequence with DNA sequence databases revealed the presence of similar sequences immediately downstream of the 5' inverted repeat of a Tc1 transposon-like element from Atlantic salmon (Radice, Bugaj and Emmons, unpub. data, GenBank Acc. #L12206) (see Fig. 6.8). This Tc1-like transposon is markedly similar (93% over the 972 bp for which the sequence is available) to the Tc1-like sequence inserted upstream from the anticipated transcription start site of an Atlantic salmon ependymin gene (Müller-Schmid et al. 1992; see Section 5.2). The main difference between the Tc1-like sequences of the ependymin gene and #L12206 is an insertion of 1424 bp into the 5' terminus of the #L12206 sequence. It is within this insert that the sequence showing similarity to the ASNHE consensus is found. The location of the 5' end of the Tc1-like element, as identified in this GenBank entry, appears to be incorrect. The reason for believing this is that the inverted repeat, as shown in Fig. 6.8, is duplicated at the 3' end (Fig. 6.3) and it also shows homology with the inverted repeats of two other Tc1-like elements from Atlantic salmon isolated by the above researchers (GenBank Acc. #L12207 and #L12208, and Fig. 5.4) and of SALT1-SSal1 (Section 5.2). Precedents exist for tandem repeats contained within inverted repeats or within the body of mobile elements (Martienssen and Baulcombe 1989, Hankeln et al. 1994, and references therein).

There is one complete copy and part of another ASNHE-related sequence within the 1424-bp insert of #L12206. The first copy is about 70% identical to the ASNHE consensus sequence. Adjoining this first copy is part of a second ASNHE-related sequence that is 67% identical over the first 150 bp; from there the similarity ends.

Fig. 6.8. Comparison of the sequences of part of the Tc1 transposon-like sequence identified by Radice et al. (1994)(SmoTc1lt, GenBank Acc. #L12206), the Tc1-like element upstream of the Atlantic salmon ependymin gene (Ssepd; Müller-Schmid et al. 1992), and two copies of the ASNHE consensus sequence. The reverse complement sequences of the two Tc1-like elements are shown. Dots indicate identical bases and dashes indicate deletions. The two inverted repeats flanking the 1424-bp fragment inserted into SmoTc1lt (A and B) and the inverted repeat at the 5' end of the Tc1-like element (C) are underlined. The 6-bp palindrome sequence (GTTAAC) at the point of insertion has been underlined a second time.

The 1424-bp inserted fragment is flanked by 30-bp imperfect inverted repeats having the palindromic sequence GTTAAC at their terminal ends. GTTAAC is the recognition sequence for *HpaI* and when Atlantic salmon DNA is reacted with this enzyme it yields a band of approximately 1470 bp (Fig. 4.1) that hybridized to ASNHE6 (Fig. 6.9). Only one copy of this palindrome is found at the corresponding site in the Tc1-like sequence in the salmon ependymin gene (Fig. 6.8). Short inverted terminal repeats and duplication of the sequence at the point of insertion are characteristics of Class II transposable elements. No sequence similarity was found between the 1424-bp insert and any known Class II element. Furthermore, it seems unlikely that this sequence encodes a transposase usually associated with these transposons as the longest open reading frame is less than 60 amino acids. If the 1424-bp sequence is a relic of a once active transposon, and a member of the ASNHE family of repeats happened to become inserted in this transposon when it was active, this could account for the expansion and spread of these sequences in the ancestor of the Atlantic salmon and brown trout.

6.6 Summary

(1) The genomes of the two species in the genus *Salmo* contain a 380-bp repetitive element flanked by the recognition sequence of *NheI*. Repeat copies comprise more than 1.2% of the Atlantic salmon genome and do not exist in long tandem arrays as is typical of maxi-satellite DNA.

(2) A comparison of the sequences of 16 salmon and 7 trout elements revealed that subfamily structure exists and there is evidence that members of the same

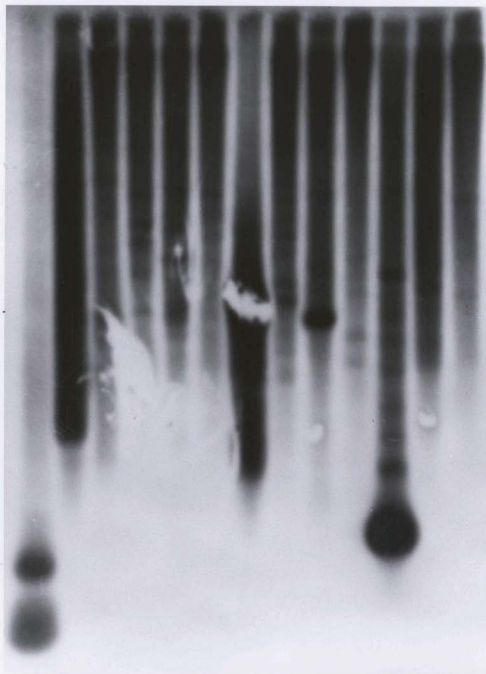
Fig. 6.9. Southern blot analysis of Atlantic salmon DNA digested with the same enzymes as shown in Fig. 4.1 and probed with the insert of BTNHE68. No specific hybridization has occurred to the bright bands evident in Fig. 4.1, with the exception of the 380-bp band of *NheI* and the 1470-bp band of *HpaI* (shown by arrows).

AluI
 AvaII
 BglI
 BglII
 BstEII
 EcoRV
 HaeIII
 HindIII
 HpaI
 HpaII
 NheI
 PstI
 SacI

bp

1470 →

380 →



subfamilies are found in both species.

(3) A search of the GenBank database indicated that sequences homologous to the *NheI* repeat are located within a 1424-bp segment inserted near the 5' end of a *TcI*-like sequence from Atlantic salmon (Radice et al. 1994). The segment is flanked by inverted repeats and there appears to have been a target site duplication at the point of insertion.

CHAPTER 7. MINISATELLITE FAMILIES - RESULTS AND DISCUSSION

7.1 Introduction

Due to the high degree of variability of VNTR loci and their potential value as genetic markers for stock and individual identification and for gene mapping (see Section 1.5), I sought to identify representatives of mini- and microsatellite repeat families from Atlantic salmon. These were isolated by means not dependent upon the use of mammalian minisatellite DNA probes. Some were obtained while isolating fluorescent-bright bands as described in previous chapters. While a majority of clones obtained from the bands included either monomer units of tandemly repeated DNA, or sequence fragments internal to interspersed repetitive elements, a minority of recombinant clones contained unrelated background DNA. Some of this DNA included minisatellite arrays and simple sequence (microsatellite) repeats. The present study represents the first detailed study of different families of minisatellite-like repeats in a lower vertebrate.

To further investigate these interspersed tandem repeat loci, a total of 1620 plaques of the Atlantic salmon λ EMBL3 genomic DNA library were screened with the salmon DNA insert of a previously isolated minisatellite clone, SsPstIL.26 (Fig. 7.1E). Following 6 days of exposure, hybridization to almost 700 (43%) of the plaques was detectable. Assuming the length of a phage insert is 15 kb and that there was only one locus per insert, this would indicate approximately 85,000 minisatellite loci related in sequence to SsPstIL.26 per haploid genome. It is possible, however, that the probe is detecting some similar sequence in otherwise divergent repeats.

Fig. 7.1. Sequences of minisatellite arrays isolated from Atlantic salmon. Dots indicate that the sequence is identical to that of the consensus (in bold), and dashes indicate deletions. Direct repeats are shown above the consensus sequences and restriction endonuclease recognition sites are indicated. Palindromes are underlined. Sequence was not obtainable for stretches of SsBgIIIU.20 and SsSaclU.44. GenBank accession numbers are shown beside each sequence name. One clone, SsBgIIIL.6, is a dimeric repeat, its two units aligned here, and designated as 6a and 6b (1D). A homologous clone, SsBgIIIL.27, contains a single monomer unit.

Next, duplicate plaque membranes were rescreened with radiolabelled poly(dA-dC)poly(dG-dT) DNA (Pharmacia), but positive plaques were much fainter and difficult to discern following a similar exposure time. Nevertheless, 46 recombinant plaques were detected, and 24 of these had previously been detected by the SsPstII.26 probe. DNA from 14 of these 24 phage clones was isolated and confirmed to be positive for both probes by Southern hybridization analysis. The DNA was next digested with *AluI* or *Pall* and the fragments subcloned into pUC18. In this way, four repeat loci were identified and sequenced: two (SsAluI16.62, Fig. 7.2D and SsAluI11.37, Fig. 7.2H) were identified by both probes, and two (SsAluI17.5, Fig. 7.2A and SsHaeIII14.20, Fig. 7.2C) only by the poly(dA-dC)poly(dG-dT) DNA.

7.2 The Isolation of microsatellite arrays

It has been estimated that the Atlantic salmon haploid genome contains 31,000 (Slettan et al. 1993) to 100,000 (Hamada et al. 1982) (GT)_n microsatellite loci, assuming a haploid genome size of 3×10⁹ bp (Hinegardner and Rosen 1972). Estoup et al. (1993) estimated 109,000 copies to be present in the genome of brown trout. The relative abundances of different simple sequence repeat motifs in fish are not known. Dinucleotide arrays isolated in the present study were either of the form (CA)_n(GT)_n or (GA)_n(CT)_n. Surveying the EMBL and GenBank databases, Beckmann and Weber (1992) also found (CA)_n(GT)_n, followed by (GA)_n(CT)_n sequences, to be most prevalent in rats. (For humans, however, (A)_n(GT)_n and (CA)_n(GT)_n arrays were the most common).

Fig. 7.2. Sequences of repetitive elements isolated from Atlantic salmon. Microsatellite arrays are shown in bold. Identical nucleotides are shown by "|", and direct repeats are indicated by a line above the sequence. Palindromes are underlined. For SsBgIIIM.26 (2G) inverted repeats are shown above the sequence. Sequence was not obtained for internal sections of SsSacIU.24 (2B) and SsBgIIIU.48 (2E).

Imperfect arrays of tetranucleotide motifs were also detected in Atlantic salmon. Short clusters of $(GACA)_n/(GGCA)_n$ are in the 3' region of the complementary strand of SsAluI16.62 (Fig. 7.2D). Clone SsStuI6.62 (Fig. 7.3A) contains a degenerate array of the $(AGCT)_n$ quartet close by a pyrimidine-rich region. An analogous repeat environment exists in a stretch of DNA downstream of the *S. salar* growth hormone gene I (Fig 7.3B, GenBank Acc. #X61938; R. Male et. al., unpub.).

7.3 Different minisatellite families

The sequences shown in Fig. 7.1 are internal portions of longer tandem arrays. Their cloning was made possible by fortuitous mutations in some monomer units. On the other hand, sequences presented in Figs. 7.2 to 7.4 include short tandem arrays together with flanking sequence. It is possible that these are precursor units which could eventually be expanded into longer arrays through mechanisms of replication slippage, amplification or unequal crossing-over. Certain sequence features suggest modes of evolution of these tandem arrays.

Repeat units of the tandem arrays range in size from 28 nt (SsPstIL.26, Fig. 7.1E) to 59 nt (SsBgIIIU.48, Fig. 7.2E). A search of the EMBL and GenBank databases using FASTA (Pearson and Lipman 1988) and the BLAST network service (Altschul et al. 1990) failed to find sequences with significant similarity to any of the cloned repeats. The repeats were examined for the presence of core sequences characteristic of many other eukaryote minisatellites. The presence of a core region (GGGCAGGAXG) within minisatellite monomer units having similarity to the *E. coli*

Fig. 7.3. Two similar repetitive elements isolated from Atlantic salmon and containing degenerate (ACGT)_n tetranucleotide and pyrimidine dinucleotide repeats (shown in bold). A) SsStuI6.62, sequence from the 3' end of a cloned *StuI* fragment containing a portion of the SALT1-SSaII transposon-like element (see Section 3.3.3). B) SSGGH, sequence downstream of the 3' end of the *S. salar* growth hormone gene I (GenBank Acc. #X61938; R. Male et al., unpub).

A) *SoSlui6.62*

```

AGGCC TAAAAACACAACAGCTAACGTTTGCTAGCTAGCTAGCTAAACCACAGTTAGCCAGAGCAATATGGCTAGGTAG
CTGGAAAAATCTAGCTAGATAGCTAACTTCCTATGCTAAATCATCCATCAGAATGAATGTATCCCCCGAAAAA AAC
TAAACAGATGGCATGGAAAACTTACTTCTCTCTCTGCTGTGACCTTTTGTCATCTACAACACTTACTTCTTTCCG
GCCAATCTATCATGTCTGGATTTCACACAGTCTCTCTCGACCCATCTTTACAAGTGACATTGGAGGTCATCGTTATTTC
AAGCGTGTGTGACAATTTGTTTAAATAACAGCACATTGAAATGTAGTTTATACATAAAGTAAAAATAATTGTCTCTAT
GTACTGCAACATTCATTCTATTTAAAAAAGAAAGTATTATGCTGCAGCTATNAACCCTGACCTGTTCAACCGGACGTTG
TACTG---TCCAGACCTG
|||
CTG---TCCAGACCTG
|||
CTGGGAACCTGACCTGTTTACCCGGACAGTGCTACTGTGCCAGACTGCTGTTTNCACNCTCTAGAGACAGCAGGAGC
GTAGAGATA

```

B) *SSGH - S. salar growth hormone gene I* (R. Male et al., unpub; Acc. #X61938)

```

GCATTAAACATTTATGCAATTTAGCTAGCTAGCTGCAGTTGCTAGCTAATTTCCAATTTAGCTAGCTTGCTTGCT
AGCTAATTTGGGATATAAATGTTGAGTTGTTATTTTACCTGAATTCACAAGGTCCTCTACTGACAATTAATCCACAC
ATAAACAGCTCAACCGAATCGTTTCAGTCATCTCTCCATCCAGGCTTTTCTTCTTTGGATTCTATATGGCATTGG
CAGCTAACTTTCATAGTATTGCCACGACGACCGACCTACCTCAGTTTCATCTTCAATCACC
CACGTGGTAT-----AACCAATGCGGATGG
|||
CACGTGGGTATCTGCTTCTATAAACCAATGAGGAGATGGGAAAGGCAGGACTGTACCGTATTCAACGTCACAAATAGAA
CTAACCTCTATTTAGGGCTTGGCAACGCAGATGCTCGTTGGCGAGCGCAGCAGTGTGGTGCAATAATTGAATAACAT

```

Fig. 7.4. Sequence of clone SsPstIL.48, isolated from Atlantic salmon. Other similar sequences shown are: Tes1, a portion of a Tc1 transposon-like element from hagfish (Heierhorst et al. 1992); SALT1-SSal1 (see section 5.1); sequence immediately 3' of a SINE element present in the genome of coho salmon (Hpa(OK)-51), but absent from the same locus in kokanee salmon (Hpa(ON)-51)(Murata et al. 1993); *Bufo bufo* B13.2, sequence downstream of a microsatellite locus in European toad (GenBank Acc. #U05292, Scribner et al. 1994).

*SsPst*IL.48 GTGAATGACCTTTCACTCAGTCTCTCTGTCTCTCATTCGATATCTCTCTTTCTCACCTCTCTCTCC
*SsPst*IL.48 ATTCCACACTCTATCCTTTTCTCTCCCTCTCTCTTGNSCCTAATATATCTATT7TCTCCCTGTCTCTCTCC
*SsPst*IL.48 CAATCTCATCCCAATAATCTCTCTACCGGCATTCAGGTGGT7TCTGTC

*SsPst*IL.48 AGAAC**CAGTACAGAGATAGCT**TGTAGAACAATCAGGTGGT**TTCTGTT**

*SsPst*IL.48 AGAAC**CAGTAGAGAGATAGCT**TGTAGAACAATCAGGTGGT**TTCTGTT**

*SsPst*IL.48 AGAAC**CAGTAGAGAGAGAGCT**TGTAGAATAATAGTGAAGACATCAAAGTATGAATAACACATATGAA

Hpa(ON) - 51/Hpa(OK) - 51
 AGAAGAAATATGAAGACATCAAAACTRTGAATTAACACATATGGA
 TCTATGACTA-TGAAGGCATCAAAACTATGAATTAACACATGTGGA

Bta1.2

TES1 TAGGGACCGATGAATCAAAATTTGAJAATCTTCGT-----TTCA'TCACGAGGGTTTTGTACGCTGTG

*SsPst*IL.48 TGTCAGTGATGAGTCCAATTTGAGATTTTGG-----TTCCAACCACCGTGCTTTGTGAGACGAG

*SALT*I-*SsSal*i TCTGGTCTGACGA AACCAATATAGAACTCTTGGCCTGAATGCCAAGCATCAGCTTT-----

TES1 AGTAGGCAAAAGGATAGTTCTCTCATTTGTTGACACAACTGTCAAACATGGAGGAGGAAGCTT'GATGATC

*SsPst*IL.48 AGTAGGTGAACGGATGATCTTCGATGTGTGGTCCCACTGTGAAGCAAGGAGGAGAGGTGT'GATGGTG

*SALT*I-*SsSal*i -GGAGGAACCTG-----GCACCATCCCACTGTGAAGCATGGTGGTGACAGCATCATGCTG

TES1 TGGGGGCTCTTCGCTGGATCCAGAGTCCGG--CGACTTGCACAGAGTGACTGGCACCTGAAACCGAAAGGG

*SsPst*IL.48 TGGGGGTCTTTCTGCTGGTACACTGTCTG--TGATTATTTTAGAATTC AAGCGAGGCTTAACCGAGCATGG

*SALT*I-*SsSal*i TGGTTATGTTTTCAT-GCGCGAGGAGCTGGATGACT-AGCCAGGATCGAGGGAAAAGATGAACGG---AGC

TES1 CTACCACAGCATTTTGACGCGGCATGCAATACCCCTCTGGTCTACG

*SsPst*IL.48 CTACCACAGCATCTCGAC

*SALT*I-*SsSal*i AAAGTACAGAGATCCTTGATGAAACCTGTCTCAGGACACTCA

"Chi" recombination signal prompted Jeffreys et al. (1985a) to propose that these repeats are "hotspots" for recombination. As summarized by Wright (1993), the motif C/GAGG is shared by many of the minisatellites isolated to date. As the original sequences (33.6 and 33.15) of Jeffreys et al. (1985a) were most often used as probes for their isolation this is to be expected. The C/GAGG motif is present in SsAluI16.70, SsBglIIL, SsPstIL.26, SsAluI17.5, SsAluI16.62, SsHpaII.24, and SsPstIL.48. A comparison of the consensus sequences of the other Atlantic salmon tandem repeat sequences does not reveal a common core region similar to that of Jeffreys et al. (1985a) or similar to other types of minisatellites, including those with (AT)-rich units, which have been isolated from human and murine genomes (see Vogt (1990) for review). I conclude that some of the salmon sequences represent previously uncharacterized repeat families.

All repeat arrays contain internal short direct repeats 2 to 6 nt in length. Deletions within the monomer units of arrays shown in Fig. 7.1 tend to be localized at these short repeats. Short direct repeats have been associated with deletions in a variety of organisms, both prokaryote and eukaryote (Thacker et al. 1992; Trinh and Sinden 1993, and references therein). The presence of short tandem repeats within longer repeats such as SsSacIU.44, SsBglIIL.6, and SsPstIL.26 also implicates slipped-strand mispairing as a mechanism important in expansion of their unit sequence (Levinson and Gutman 1987).

7.4 The junctions of minisatellite arrays - association with other repeat types

The fact that SsPstIL26 and poly(dA-dC)poly(dG-dT) DNA cohybridize to recombinant phage of the salmon genomic λ EMBL3 library, together with sequencing results presented in Figs. 7.2 to 7.4, suggests that microsatellite dinucleotide arrays are associated with short, tandem repeats at many loci in salmon. Bentzen and Wright (1993) have also identified a cryptic microsatellite present near the 3' end of a minisatellite array in Atlantic salmon.

Evidence for the association of microsatellites with minisatellite sequences in other organisms is anecdotal and involves, for the most part, minisatellite repeat units harbouring short stretches of simple sequence. One family of interspersed tandem repeats in cattle possesses a 29-bp core sequence ending in variable (GT)_n arrays (Kashi et al. 1990), a phenomenon analogous to that of the SsAluI17.5 element (Fig. 7.2A). Human VNTR alleles with internal di- and tri-nucleotide arrays have been characterised (Armour et al. 1992; Iwasaki et al. 1992).

Levinson and Gutman (1987) propose a process of rounds of slipped-strand mispairing for the initial generation of microsatellites and then subsequent rounds of SSM interspersed with mutational events for the production of minisatellites. It has also been suggested that unequal crossing-over plays a role in the evolution of longer tandem repeat motifs from simpler short ones when a minisatellite tandem array has reached a particular size. Wright (1994) has also presented a case for microsatellites as the progenitors of minisatellites. The presence of simple sequence motifs within the

monomer units of minisatellite arrays would support this scenario. It is not unreasonable to suggest that the repeat units of SsAlu17.5 (Fig. 7.2A) or SsSaeI.U.24 (Fig. 7.2B) have been amplified from d(GT)d(CA) dinucleotides. It is less clear that the long microsatellite arrays shown in Figures 7.2B-D and Fig. 7.4 have been involved in the birth of their minisatellite-like neighbours. In this case, one might expect to see a zone of transition through which microrepeats are gradually expanded. Rather, I failed to detect simple sequence arrays contiguous with minisatellite repeats, the two being separated instead by unique sequence. While SSM may be involved in the generation of short sequence repeats, recombination may act to bring unrelated repeat types into juxtaposition. The power of d(GT)_n dinucleotide tracts to promote homologous recombination through the ability to form left-handed Z-form DNA has been demonstrated (Trecó and Arnheim 1986; Wahls et al. 1990).

Furthermore, in Figures 7.2 to 7.4, palindromes and quasipalindromes (repeats separated by a short spacer) tend to be found near the junctions of tandem arrays and their flanking DNA. The minisatellite array from Atlantic salmon described by Bentzen and Wright (1993) also has imperfect inverted repeats near its 5' (CGACCAGAAGGTCG) and 3' (AGAAAACCTACACCTAGCAGTTGTCT) ends capable of forming cruciform structures. Palindromes have been variously associated with increased recombination, DNA amplification, and local sequence excision (reviewed in Reed et al. 1994). Other studies have detected palindromes at sites of recombination in satellite DNA (Kiyama et al. 1987; Plohl and Ugarković 1994). It has been proposed that these could form cruciform structures which might be involved in

recognition by some recombination enzyme (Hyrien et al. 1987, and references therein).

7.5 Characterization of SsPstIL.48

In humans it is not unusual to find microsatellite and minisatellite sequences adjacent to other interspersed repeats (Das et al. 1987; Armour et al. 1989, Rogaev 1990). By searching the DNA databases, Beckmann and Weber (1992) identified many human microsatellite sequences close to *Alu* repeats. In the case of Atlantic salmon, several repeat types were found to be juxtaposed at the SsPstIL.48 locus (Fig. 7.4).

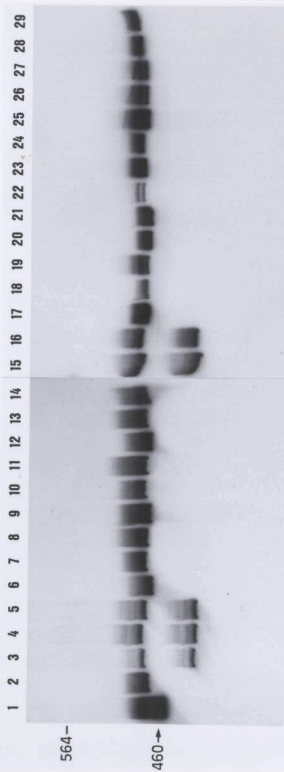
As discussed in Chapter 5, digestion of Atlantic salmon genomic DNA with *Pst*I produces two visible bands of 580 bp and 800 bp following electrophoresis and staining with ethidium bromide. DNA in these bands corresponds to regions internal to the Tc1 transposon-like SALT1 element. Clone SsPstIL.48 (Fig. 7.4) was isolated from the 580-bp band. A search of the GenBank database revealed that 220 bp at its 3' end showed greatest sequence identity with a portion of the Tc1-like element (Tes1) isolated from Pacific hagfish (65%; Heierhorst et al. 1992) and to lesser degree with the *Uhu* transposon of *Drosophila* (54%; Brezinsky et al 1990). Similarity with the SALT1-SSal1 homologous sequence was 58 percent.

Unrelated sequence appears to have been inserted into the coding region of SALT1-related sequence of SsPstIL.48, perhaps by means of recombination between non-orthologous repeat loci. The 5' end contains an imperfect (CT)_n microsatellite followed by three 48 bp tandem repeats. A quasipalindrome exists at the junction of these two repeat types.

Unexpectedly, a 43-nt stretch between the last tandem repeat and the start of the SALT1 homologous region of SsPstIL.48 is 95% identical with sequence immediately 3', but not part of a SINE element (Hpa OK-51) of coho salmon (*Oncorhynchus kisutch*). (An orthologous locus, Hpa ON-51, in kokanee salmon (*O. nerka*) is missing the SINE (Murata et al. 1993)). There also exists close similarity with a region 65 bp from the start of a (GT)_n microsatellite locus in the European toad (*Bufo bufo*; Scribner et al. 1994; GenBank Acc. #U05292). No explanation can be offered for the convergence of these three taxonomically diverse sequences.

To investigate intraspecific variation in the SsPstIL.48 locus, two oligonucleotide primers were synthesized, one upstream of the 5' microsatellite, and the other contained within the SALT1 homologous region (Table 3.1 and Fig. 7.4). The latter primer was end-labelled with [γ -³²P]ATP, and both primers were used for PCR amplification of DNA from salmon of Canada, Ireland and Scotland (see Section 3.4.3). Dimeric bands differing in size by only a few nucleotides were resolved in some individuals (Fig. 7.5). These individuals may be heterozygous for this locus, having lost one or two dinucleotide repeats from a single allele. Single bands differing in size by more than 30 nt were also generated for a few individuals, heterozygotes with longer allelic deletions. In all cases the upper bands were longer than the PCR product generated by amplification of plasmid DNA (Fig. 7.5, lane 1). It is unclear if this discrepancy is due to a deletion artifact generated during cloning of SsPstIL.48.

Fig. 7.5. PCR amplified alleles of the SsPstIL.48 locus in salmon of Newfoundland, Ireland, and Scotland. SsPstIL.48 plasmid DNA: lane 1 (and size indicated by an arrow). Atlantic salmon: lanes 2 to 4, Hatchery Brook, Newfoundland; lane 5, Conne R., Newfoundland; lanes 6 to 8, LaHave, Nova Scotia; lanes 9, 10, Black Brook, Newfoundland; lanes 11, 12, Little R., Newfoundland; lanes 13, 14, Grand Codroy R., Newfoundland; lanes 15 to 17, Gambo R., Newfoundland; lanes 18 to 22, Scotland; lane 23, Owentogher R., Donegal, Ireland; lanes 24, 25, Ray R., Donegal, Ireland; lanes 26 to 28, Swilley R., Donegal, Ireland; lane 29, L. Melvin, Ireland. The 564-bp marker shown was derived from the comigration of end-labelled *Hind*III digested λ DNA. Under these PCR conditions (see Section 3.4.2) no amplification products were detected for brown trout, masu salmon, Arctic charr, chum salmon, grayling, whitefish, sea lamprey, shark, skate, sturgeon, winter flounder, *Rana* sp., mouse, or human DNA, nor were products detected for a control reaction from which DNA was excluded (not shown).



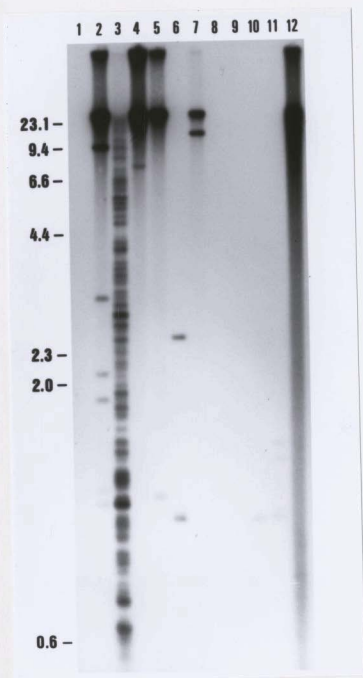
7.6 Phylogenetic Distribution of the Minisatellite Sequences

To determine the phylogenetic distribution of the minisatellite elements shown in Fig. 7.1 and so provide evidence for their relative times of amplification and their organization within salmonid genomes, restriction enzyme digests of genomic DNA from salmonid and other fish species, together with *Rana* sp., bird and human were subjected to Southern blot analysis under low stringency conditions.

In Fig. 7.6A, Atlantic salmon DNA was digested with different restriction endonucleases and probed with the insert of SsBgIIIU.20. *Pst*I completely digests each array at a site internal to the monomer units (lane 1); these have run off the gel and no hybridization is seen. *Sau*3AI, on the other hand, cut infrequently within the repeat array generating a multi-banded fingerprint pattern (lane 3). When digested with *Alu*I, which does not cut within the repeat units of SsBgIIIU.20, a small number of bands were detected (lane 5).

Hybridization bands for SsBgIIIU.20 were found for other species of the Salmoninae subfamily, including brown trout, Arctic charr, rainbow trout, masu salmon, and chum salmon (the latter appearing as an intense smear), but not Japanese charr or brook charr. (Small molecular weight fragments from the latter two species could have been lost from the bottom of the gel). SsBgIIIU.20 repetitive DNA was not detected in whitefish or Arctic grayling and must, therefore, have appeared subsequent to the divergence of the Salmoninae subfamily from the rest of the Salmonidae species, probably prior to the Pliocene period. Fossil records for Salmonidae in general are sparse, but a fossil of the genus *Salvelinus* dating from 10 million years ago (Ma) has

Fig. 7.6. Zoo blot analysis for SsBglIIU.20. Lanes 1 to 5, Atlantic salmon, digested with *Pall*, *RsaI*, *Sau3A*, *SacI*, and *AluI*; 6, brown trout; 7, Arctic charr; 8, Japanese charr; 9, brook charr; 10, rainbow trout; 11, masu salmon; 12 chum salmon. DNA of lanes 5 to 12 was digested with *AluI*. Other species tested and showing no DNA hybridization included sea lamprey, dogfish shark, skate, Atlantic sturgeon, Atlantic cod, winter flounder, *Rana* sp., gannet, and human. Hybridization was at low stringency.



been identified (Smith et al. 1982). The *Oncorhynchus* lineage arose more than 6 Ma (Behnke 1992; Stearley and Smith 1993).

A typical multi-banded fingerprint pattern was produced when the insert of SsBglIIL.6 was hybridized to *S. salar* DNA cut with *AluI* (Fig. 7.7, lane 2); *PaII* digestion yielded strongly hybridizing bands of low molecular weight (lane 1). The presence of these two restriction sites in some SsBglIIL.6 repeats (see Fig. 7.1D) would tend to increase the number of hybridization bands by dissecting repeat arrays. Faint hybridization bands were also detected for Arctic charr, sea lamprey, Atlantic cod, and winter flounder, but the probe may be detecting unrelated repeats having some internal sequence similarity in these distantly related fish species.

An oligonucleotide probe based on the sequence of SsSacIU.44 strongly hybridized to Atlantic salmon high molecular weight DNA remaining in the well and to a *PaII* fragment greater than 23 kb in length (Fig. 7.8, lane 1). Several fainter and shorter fragments were also seen both for Atlantic salmon and some other species of the subfamily Salmoninae. The continued presence of high molecular weight bands using other restriction endonucleases confirmed that SsSacIU.44 exists in very long tandem arrays at a small number of loci (Fig. 7.9). *AluI* appeared to cut all monomer units; no probe hybridization was detected when genomic DNA was cut with this enzyme (data not shown). Hybridization to DNA of brown trout was not detected, although very faint bands were seen for other species of the Salmonidae, including whitefish. Therefore, amplification of both SsBglIIL.6 and SsSacIU.44 repeat sequences to high copy number and long arrays has occurred within the genome of *S.*

Fig. 7.7. Zoo blot analysis of the distribution of SsBglIII.6 in fish and other vertebrate species. DNA was digested with *Pst*I (lane 1) and *Alu*I (lanes 2-19). Lanes 1, Atlantic salmon; 2, Atlantic salmon; 3, brown trout; 4, chum salmon; 5, masu salmon; 6, rainbow trout; 7, Arctic charr; 8, grayling; 9, whitefish; 10, sea lamprey; 11, dogfish shark; 12, skate; 13, Atlantic sturgeon; 14, Atlantic cod; 15, winter flounder; 16, *Rana* sp.; 17, gannet; 18, mouse; 19, human.

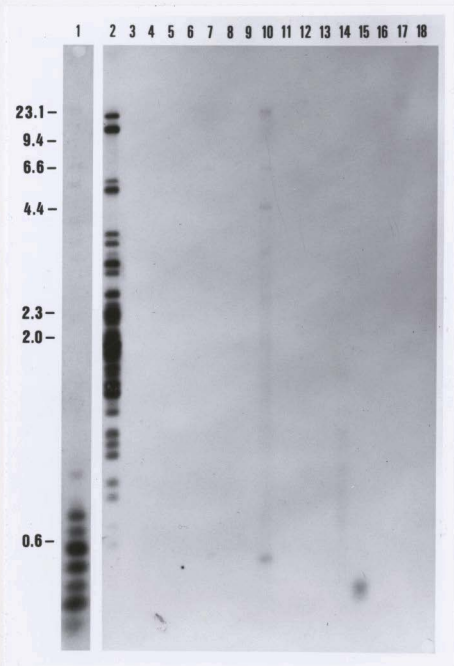


Fig. 7.8. Zoo blot analysis for SsSacIU.44 oligonucleotide probe (see Table 3.1). Hybridization was performed under low stringency conditions (see Section 3.5). Only species of the subfamily Salmoninae are shown. Lanes 1, Atlantic salmon; 2, brown trout; 3, Arctic charr; 4, Japanese charr; 5, brook charr; 6, rainbow trout; 7, masu salmon; 8, chum salmon; 9, Arctic grayling; 10, whitefish. Other species tested and showing no DNA hybridization included sea lamprey, dogfish shark, skate, Atlantic sturgeon, Atlantic cod, winter flounder, *Rana* sp., and human. All DNA was digested with *PvuII*.

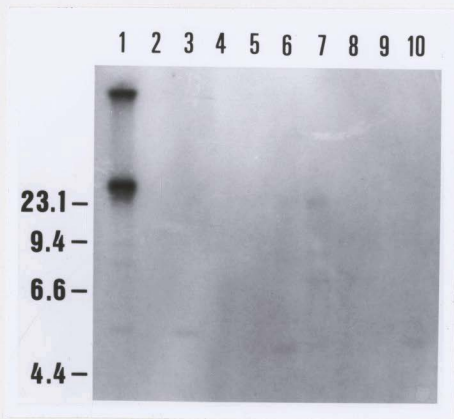
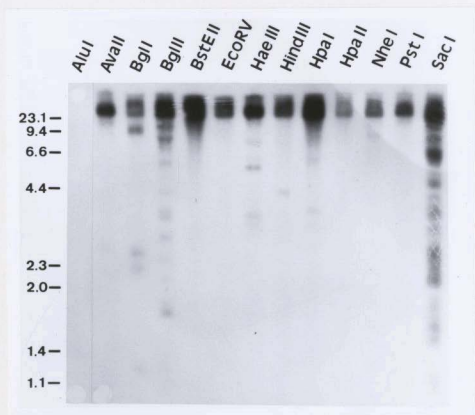


Fig. 7.9. Southern blot analysis of Atlantic salmon DNA digested with the same enzymes as in Fig. 4.1 and probed with the oligonucleotide probe of SsSacIU44. Hybridization was performed under high stringency conditions using the modified DNA hybridization protocol of the RCMP, Ottawa (conditions were as described in Section 3.6. Washing was at 60°C). Because *AluI* cut within each monomer unit no probe hybridization was detected in lane 1.



salar following its divergence from that of *S. trutta*.

The insert from clone SsAlu16.70 yielded distinct hybridization bands for all charr, trout and salmon species tested (Fig. 7.10). Detection in whitefish suggests the presence of this repeat prior to the Salmoninae subfamily split. *Eosalmo* of the Eocene period (45 Ma) is the most primitive salmonine known (Wilson 1974; Stearley and Smith 1993). Hybridization to DNA of non-salmonid appeared as smears without discernible banding; it is not certain if the signal derived from non-specific binding.

SsPstfL.26 (Fig. 7.11) generated strong multi-locus fingerprints in all Salmoninae species; following long exposure of 6 days, hybridization was detected for almost all other species tested. However, significantly weaker signal indicates considerable sequence divergence of the repeat loci being detected in the non-Salmoninae lineages. Among the greater than 35 bands visible for Atlantic salmon, there is considerable intrapopulation VNTR variation, even under hybridization conditions of high stringency (Fig. 7.12B, and data not shown). This repeat sequence, therefore, promises to be valuable as a general multilocus fingerprinting probe for salmon and charr species.

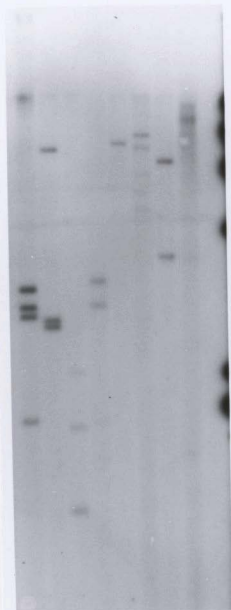
7.7 Use of SsBglIIU.20 as a probe for the analysis of population structure

Under high stringency hybridization conditions, BglIIU.20 will not hybridize to non-*Salmo* species. Furthermore, in brown trout only two or fewer fragments are detected, each less than 3 kb in size (Fig. 7.12A, lanes 10 and 11, and data not shown), indicating selective amplification of this repeat sequence to higher copy number and longer arrays within the genome of *S. salar* following its divergence from *S. trutta*.

Fig. 7.10. Zoo blot analysis for SsAluI16.70. Lanes 1, Atlantic salmon; 2, brown trout; 3, Arctic charr; 4, Japanese charr; 5, brook charr; 6, rainbow trout; 7, masu salmon; 8, chum salmon; 9, grayling; 10, whitefish; 11, sea lamprey; 12, dogfish shark; 13, skate; 14, Atlantic sturgeon; 15, Atlantic cod; 16, winter flounder; 17, *Rana* sp.; 18, human. All DNA was digested with *PvuII*. Although no hybridization is apparent for grayling (lane 9), it should be noted that comparatively less DNA was used. The markers shown are λ DNA cut with *HindIII* and 32 P-labelled.

1 2 3 4 5 6 7 8 9

10 11 12 13 14 15 16 17 18



23.1

9.4

6.6

4.4

2.3

2.0

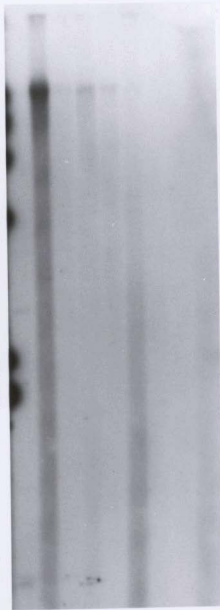


Fig. 7.11. Zoo blot analysis for SsPstIL.26. Lanes 1, Japanese charr; 2, Atlantic salmon; 3, brown trout; 4, Arctic charr; 5, brook charr; 6, rainbow trout; 7, masu salmon; 8, chum salmon; 9, grayling; 10, whitefish; 11, sea lamprey; 12, dogfish shark; 13, skate; 14, Atlantic sturgeon; 15, winter flounder; 16, *Rana* sp.; 17, gannet; 18, human; 19, Atlantic cod. All DNA was digested with *Pst*I. Hybridization was at low stringency. No hybridization is seen for gannet (lane 17), but the DNA from this sample was degraded.

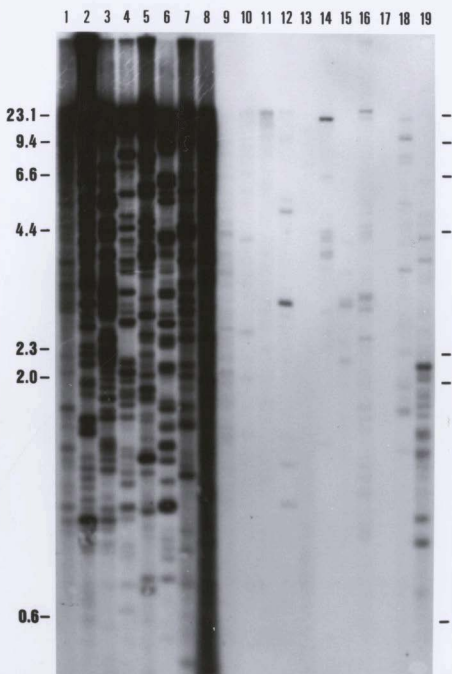
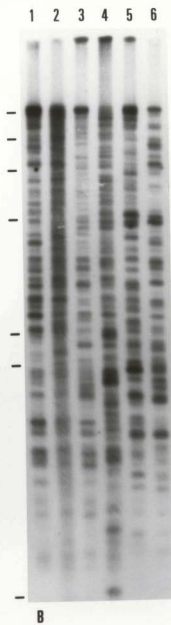
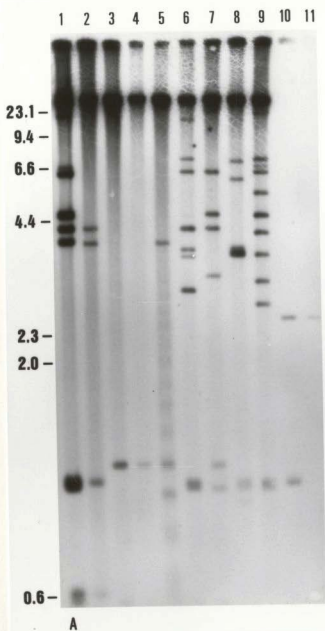


Fig. 7.12. Analysis of genomic DNA from Atlantic salmon of Eastern Canada and Ireland. A) DNA was digested with *AluI* and probed with SsBgIIIU.20. Atlantic salmon: lane 1, Little R., Newfoundland; lane 2, Grand Codroy R., Newfoundland; lanes 3 to 5, LaHave R., Nova Scotia; lanes 6 to 7, Owentogher R., Donegal, Ireland; lane 8, Ray R., Donegal; lane 9, Swilley R., Donegal; lane 10, brown trout, Newfoundland; lane 11, brown trout, Ireland. B) DNA of Atlantic salmon digested with *PaeI* and probed with SsPstIL.26. Lanes 1, 2, LaHave, Nova Scotia; lane 3, Little R., Newfoundland; lane 4, Grand Codroy, Newfoundland; lane 5, Owentogher R., Donegal; lane 6, Ray R., Donegal. Hybridization was at high stringency (see Section 5.6).



The segregation of loci detected by SsBgIIIU.20 was examined in family pedigrees of female Atlantic salmon and male brown trout parents and their F1 hybrid offspring (McGowan and Davidson 1992). The results of all crosses were as predicted for Mendelian inheritance. Germline stability of the minisatellite loci was demonstrated by all crosses tested. Results for one cross are shown in Fig. 7.13. Since the SsBgIIIU.20 probe would not hybridize to the DNA of the particular brown trout individual shown here, all bands observed in the offspring are derived from the *S. salar* mother. However, strong linkage between DNA hybridization fragments detects clustering of repeat arrays at some loci.

Hybridization of the SsBgIIIU.20 probe to DNA of Atlantic salmon digested with *AluI* generates from one to over ten polymorphic bands (most greater than 2.5 kb in length, but a second tier of fainter and more diffuse bands less than 1.5 kb may be seen; Fig. 7.12A). It was considered, therefore, that this probe might detect population-level markers in a species which reveals only limited genetic variability with isozyme or mitochondrial DNA analysis (Davidson et al. 1989) and rDNA RFLPs (Cutler et al. 1991). To test this idea, DNA of 56 salmon from Ireland and 7 from Scotland, and 87 salmon from Newfoundland and 14 from mainland Canada was digested with *AluI* and subjected to Southern blot analysis under high stringency conditions. A "genotype" was assigned to each individual fish based on its overall composite multilocus banding pattern (bands less than 1.5 kb in length were ignored as difficult to score). Most of the genotypes of British salmon were individual-specific. With one exception, no Canadian genotype was found in fish of British origin (Table

Fig. 7.13. Analysis of genomic DNA from F1 offspring (lanes 2 to 13) and parents of a female Atlantic salmon (1634♀, lane 1) X male brown trout (1619♂, lane 14) hybrid cross (samples were a gift from C. McGowan). DNA was digested with *AhaI* and probed with the insert of SsBgIIU.20. Hybridization was at high stringency.



SAMPLE SITE	GENOTYPE NUMBER																	Number of Individual- Specific Genotypes
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
IRELAND																		
Donegal, Owentogher R.			1		1				2									6
Keewagh R.	1	2																2
Ray R.			1	1						1								5
Glen R.	3			1		1		2										2
Swilly R.	1					1												7
L. Melvin		2	1		1		2											4
Antrim, Glenarm R.				1														3
SCOTLAND																		
Four locations									1									5
NEWFOUNDLAND																		
East Coast, Avalon Peninsula													7	4				2
Northeast Coast, Gambo R. System									1	3		1	1					2
Black Br.										2		3						2
Indian Br.										4	1							2
South Coast, Hatchery Br.										8					3			1
Little R.										8	1	2						1
Conne R.										1		3						2
Grey R.										1		1						5
West Coast, Grand Codroy R.										13	1					2		5
MAINLAND CANADA																		
Nova Scotia, LaHave R.										1	4							1
New Brunswick										1	5							
Quebec, Ungava Bay										1	1							

1). One measure of genotypic diversity is given by the nucleon diversity index, $h = (1 - \sum(x^2))/(n)/(n-1)$, of Nei and Tajima (1981), where x is the proportion of each genotype and n is the total number of individuals. Variability was much lower for Canadian ($h = 0.78$) than British ($h = 0.98$) salmon; in fact, 72% of the fish tested were classed as having one of three genotypes. Mitochondrial DNA haplotypes also may be used to predict continent of origin for Atlantic salmon, although "European" haplotypes have been detected in Newfoundland (Bermingham et al. 1991; Birt et al. 1991; McVeigh et al. 1991). It remains to be seen if use of the SsBglIIU.20 probe will prove more sensitive than previous methods for distinguishing regional population differences. It is noteworthy, however, that the two predominant genotypes (nos. 14 and 15) of fish from the Avalon Peninsula of Newfoundland were not detected in fish from the west coast of the island or from mainland Canada. A more thorough investigation of relatedness at the individual or population levels should involve statistical analysis of shared band fractions (Lynch 1988). This was not attempted here.

7.8 Summary

(1) Minisatellite and microsatellite loci were isolated from both agarose gels and the genomic salmon library and characterized.

(2) Southern blot analysis of the phylogenetic distribution of a subset of the minisatellites indicates one sequence to be pervasive among vertebrates, others present only in Salmoninae or Salmonidae species, and two selectively amplified to higher copy numbers and longer arrays in Atlantic salmon as compared to brown trout.

(3) There is evidence for the close association of microsatellite and minisatellite arrays at many loci. Furthermore, one tandem repeat appears to have been inserted into the transposase coding region of a SALT1 element.

(4) Two of the minisatellite loci were shown to detect VNTR loci and should prove useful for finding individual or population level markers. One reveals a greater degree of genotypic variability among Irish than among North American salmon.

CHAPTER 8. CYTOGENETIC MAPPING OF REPEAT SEQUENCES

8.1 Introduction

The development in the early 1970s of high resolution banding techniques for mammalian chromosomes was critical to the rapid development of gene mapping strategies. The ability to achieve reliable structural karyotypes permitted somatic cell hybridization and, more recently, *in situ* chromosome hybridization procedures to assume prominent positions in low resolution human gene mapping. In fact, of the human autosomal loci mapped by 1990, 38% and 22% had been mapped by these procedures, respectively (McKusick 1991).

In his detailed review of fish cytogenetics, Ojima (1983) published karyotype descriptions of about 350 fish species, summarizing a data base maintained at Japan's Kwansei Gakuin University (Fujii and Ojima 1983). The application of cytogenetics to fish species has been slow due to difficulties of maintaining cells in culture, obtaining good metaphase preparations, and reliably staining chromosomes. This is exacerbated in groups such as salmonids which have large numbers of small chromosomes. Techniques able to elongate fish metaphase chromosome preparations could improve banding resolution (Liu 1986).

8.1.1 Chromosome banding

Fish chromosomes have generally proven refractive to standard high resolution chromosome banding techniques. Most published reports have dealt with C (Centromere)- or N (NOR)-banding only. Linear banding patterns, similar to the G

(Giemsa)- and R (Reverse)-banding patterns of higher vertebrates, have been reported in a very small number of fish species, including European eel (*Anguilla anguilla*; Wiberg 1983, Medrano et al. 1988, Viñas et al. 1994), *Missgurnus anguillicaudatus* and *Monopterus albus* (Liu 1986), and brown trout (Blaxhall 1983). The scarcity of G-, Q (Quinacrine)- and R-band patterns has been attributed to differences in genome compartmentalization and intermolecular heterogeneity between warm-blooded vertebrates and lower cold-blooded vertebrates. For example, fish are deficient in (G-C)-rich isochore DNA regions (Bernardi and Bernardi 1990, Holmquist 1989, Medrano et al. 1988).

Although reproducible banded karyotypes have not yet been forthcoming, salmonid chromosomes have been studied extensively (Hartley and Horne 1984) and are among the best characterized among fish species. C- and Q-banding, both of which stain constitutive heterochromatin, have been described for many individual salmonids (Phillips and Hartley 1988; Phillips and Ihssen 1986; Phillips et al. 1985; Pleyte et al. 1989). The chromosomal locations of the genes for ribosomal RNA (rDNA) that have been transcriptionally active can be determined by silver staining the NORs, whereas the GC-specific fluorochrome chromomycin A3 appears to stain rDNA irrespective of its activity (Amemiya 1986). N-banding has been used in salmonids to map the chromosomal sites of the rRNA genes (Phillips and Ihssen 1985; Mayr et al. 1986; Phillips and Hartley 1988).

Although in its infancy, the use of restriction enzymes to produce modified C-banding patterns, originally used on moray eel (*Muraena helena*; Cau et al. 1988), has

been applied to salmonids (Lloyd and Thorgaard 1988; Hartley 1987, 1991; Lozano et al. 1991; Sánchez et al. 1990; 1991; 1993) and tench (*Tinca tinca*; Padilla et al. 1993). Even more promising is replication banding, successfully used to generate banded karyotypes for the cyprinids rudd (*Scardinius erythrophthalmus*) and roach (*Rutilus*; Hellmer et al. 1991), scorpion fishes (Giles et al. 1988), and salmonids (Delany and Bloom 1984; Loranzo et al. 1991; Sánchez et al. 1993; Pendás et al. 1993a; Takeaki et al. 1993). A technique with the ability to identify unambiguously individual chromosomes would have immediate application in studies of fish gene mapping, especially if used in conjunction with *in situ* chromosome hybridization.

8.1.2 *In situ* hybridization

In situ chromosome hybridization involves the direct visualization of a cloned DNA sequence at a particular chromosomal location by hybridizing the labelled clone to preparations of metaphase chromosomes. *In situ* hybridization provides a truer picture than linkage analysis of the actual physical distances of genes along a chromosome. (This is biased somewhat because chromatin condensation is nonuniform). In addition, this procedure allows genes to be mapped for which no polymorphisms are detected. Recent developments employing nonisotopic labelled probes have improved resolution and speed over autoradiographic approaches (Boyle et al. 1992). The relative position of a sequence may be easily determined along the length of an unbanded chromosome. In combination with banding, the sequence is more precisely localized, irrespective of the state of chromatin condensation.

Furthermore, by utilizing probes differentially labelled with biotin and digoxigenin, sequence as close as 100 kb along the chromatin of interphase nuclei may now be visualized, a resolution comparable to that provided by pulsed field gel electrophoresis (Lawrence et al. 1990).

In fish, fluorescent in-situ hybridization (FISH) has been used to localize centromeric satellite DNA in *Sparus aurata* (Garrido-Ramos et al. 1994) and to confirm that a tandem repeat in Japanese hagfish is germ-line restricted (Kubota et al. 1993). More extensive work has been conducted in salmonids. Using the FISH technique, histone DNA has been shown to be tandemly repeated at a single locus in brown trout, Atlantic salmon and rainbow trout (Pendás et al. 1994a). Two-colour labelling with digoxigenin and biotin has pinpointed 5S rDNA to a heterochromatic arm adjacent to the major ribosomal DNA locus in Atlantic salmon (Pendás et al. 1993b; 1994b). In brown trout, in addition to the major locus, numerous minor NORs have also been identified in the brown trout chromosomal complement (Pendás et al. 1993c).

8.2 Results and Discussion

Previous success (Yamazaki et al. 1989; Yamazaki and Goodier 1993) at preparing chromosomes from embryos of Pacific salmonids by the method of Yamazaki et al. (1981) was not transferrable to Atlantic salmon. Excessive cell debris on the prepared slides made visualization of the chromosomes difficult. The highest mitotic indices were obtained from embryos of 10 to 20 days post-fertilization.

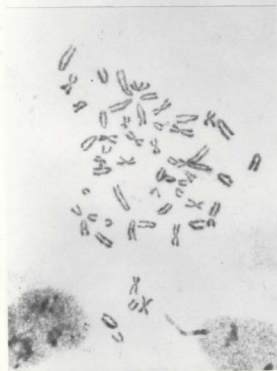
Metaphase preparations obtained from short-term Atlantic salmon blood cultures

were of better quality, chromosomes tending to be more spread out with less debris. However, mitotic index was usually low. It should be noted that blood was obtained from post-spawning, non-reconditioned fish. It is unclear if the generally poor condition of these specimens might have adversely affected lymphocyte cell division.

Attempts were made to prepare metaphase chromosomes from small-scale cell cultures of embryonic or liver tissue. Confluent cultures were easily obtained from cells of both sources. The best metaphase preparations were prepared from embryonic cell cultures which had been treated with PHA-M and passaged once. However, again results were inconsistent. The ability to obtain chromosomes decreased with frequency of subculture. Furthermore, polyploid and aneuploid cells were sometimes observed and their numbers appeared to increase with the age of the culture. The metaphase plates of Fig. 8.1 were obtained from a single culture: both diploid (A,B) and tetraploid cells (C,D) are present.

Most of the metaphase slides obtained from blood or tissue culture were used in chromosome banding experiments (Fig. 8.2). Restriction enzyme banding was obtained for *AluI* (Fig. 8.2A) and *HaeIII*, but there were no obvious reproducible patterns and banding was occasionally seen along untreated Giemsa-stained chromosomes of control slides. This was also the case for trypsin-treated G-banded chromosomes. (Enzymatic treatment of the chromosomes of Fig. 8.2B was too prolonged, resulting in excessive degradation). Chromosomes from embryos younger than 14 days are often quite elongated and so might prove most useful for enhanced chromosome banding and gene mapping by *in situ* hybridization (Fig. 8.3A,B).

Fig. 8.1. Stained metaphase chromosomes from a single primary Atlantic salmon embryonic tissue culture harvested at 26 days post-fertilization. Slides were stored for 2 months dessicated under vacuum prior to staining. The slide was pretreated with 25 units of *AluI* for 5 hr 20 min and stained with 0.3% Wright's stain (1:4 dilution with Gurr buffer, pH 6.8) for 60 sec. Both diploid and tetraploid cells are seen. A, B: 56 chromosomes; C: 112 chromosomes; D: approximately 110 chromosomes.



A C

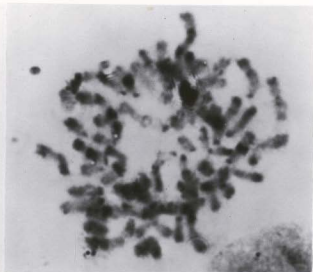


B D



Fig. 8.2. Stained metaphase chromosomes from A) a single whole embryo, 15 days after fertilization, prepared as described in section 3.9.1. The slide was treated with 15 units *Alul* as described in Fig. 8.1. B) from a single primary tissue culture of a fry 4 months post-fertilization, harvested at 24 days. The culture was treated with 0.2 $\mu\text{g/ml}$ colcemid for 19 hr before harvesting. Hypotonic treatment was with 0.8% citric acid for 25 min. In order to induce G-banding, the slide was heated at 90°C for 30 min and treated with 0.008% trypsin in Hank's Buffered Saline solution for 90 sec, followed by washing twice in 0.9% saline solution. Staining was with Giemsa for 50 sec.

A



B

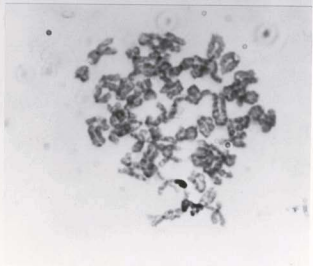


Fig. 8.3. Stained metaphase chromosomes from a single whole embryo, 14 days after fertilization. The egg was treated with 0.1 $\mu\text{g/ml}$ colcemid for 10 hr before extraction of the embryo. Hypotonic treatment was with 0.8% Na-citrate acid for 25 min. To induce G-banding the slide was treated with 0.05% trypsin for 60 sec as described in Fig. 8.2. Staining was with Giemsa for 60 sec.

A



B



However, in very young embryos, a significant amount of cytoplasm is also retained on the slide and interferes with staining and probe hybridization.

It was originally anticipated that following isolation of the repetitive probes described in previous chapters, their chromosomal distributions could be determined by biotin FISH using protocols described in Pinkel et al. (1986) and Schwarzscher-Robinson (1988). This was abandoned due to failure to obtain a consistent supply of metaphase chromosome preparations of high mitotic index. However, as cited above, work by A.M. Pendás, P. Morán, and E. García-Vázquez at the University of Oviedo, Spain has demonstrated the feasibility of using FISH to map repetitive sequences in salmonids.

CHAPTER 9. CONCLUSION

In order to identify repetitive DNA elements within the nuclear genome of Atlantic salmon, total genomic DNA was digested with restriction endonucleases and its fragments separated in agarose gels and stained with ethidium bromide. Isolation, cloning and sequencing of UV-bright bands was an efficient approach to characterizing both tandemly arrayed and interspersed repetitive DNA. Interestingly, unlike other studies employing this approach (noted in Section 1.4.1), none of the isolated bands yielded representative units of classical maxi-satellite type DNA. Elements isolated include: a) a *Bgl*I repeat element, which appears to exist in short rather than very long tandem arrays (as is characteristic of maxi-satellite DNA), some of which are in the vicinity of rDNA genes; b) members of the *Nhe*I repeat family, which also do not exist in long tandem arrays; c) interspersed elements (designated SALT1) having similarity with Tc1 transposons of nematodes; and d) various minisatellite and other short tandem repeat loci.

Some satellite DNA families may simply have been missed. Time constraints for the project necessitated that certain brightly-staining restriction fragment bands (see Fig. 4.1) be ignored. Furthermore, by separating fragments on agarose gels of rather low concentrations (0.8 to 1.2%), I selected for repetitive DNA elements of longer unit length. Those of length under 150 bp may have been poorly resolved and gone unnoticed. It is also possible that some Atlantic salmon maxi-satellite repeat units may not contain recognition sites for those endonucleases screened.

By screening a salmon genomic library constructed in λ EMBL3 vector it was

possible to isolate complete loci with unique flanking sequence for a *Bgl*I repeat array, a SALT1 element, and several micro- and minisatellite loci. However, this approach is time consuming. The maintenance of phage stocks, the need for growth to high titre, and DNA extraction protocols are tedious. Furthermore, the length of recombinant DNA inserts in phage vectors often requires that they be fragmented and subcloned into plasmid vectors prior to DNA sequencing. A more effective method for isolating highly repetitive DNA sequences involves construction of a phagemid vector library from randomly sheared and reassociated DNA (Moyzis et al. 1988). The library may be enriched for rapidly reannealing, highly repetitive DNA and avoids use of a restriction enzyme and loss of long arrays lacking its recognition sequence. DNA annealed to a C_0t of 50 may be used for screening.

9.1. Implications of repetitive DNA for species evolution

Reassociation studies (Britten and Kohne 1968; Garret et al. 1977; Schmidtke 1979; Hanham and Smith 1980) indicated that a large mass of repetitive DNA resides in the salmon genome. The present investigation of individual repeat families tends to support these observations. *Bgl*II, SALT1, and *Nhe*I repeats were estimated to comprise approximately 2.3%, 0.2 to 0.9%, and 1.2% of the genome, respectively. Minisatellite loci appear to be numerous (see Section 7.3). It has been estimated that the Atlantic salmon haploid genome contains 31,000 (Slettan et al. 1993) to 100,000 (GT)_n (Hamada et al. 1982) microsatellite loci alone. Other repeat elements found in Atlantic salmon (Moir 1988; Murata et al. 1993) may be assumed to exist in significant copy numbers.

In 1972 Ohno dismissed as "junk" DNA that part of the genome not involved in the encoding of proteins or functional RNA: a significant proportion involved repetitive sequences. Later Doolittle and Sapienza (1980) and Orgel and Crick (1980) coined the term "selfish DNA". As examples of this self-perpetuating DNA, transposable elements were regarded as molecular parasites, directing their own propagation within the genome. Far from contributing to organismal fitness, non-fortuitous insertion events could adversely affect the host organism by disrupting normal patterns of gene expression. It has been shown, for example, that mammalian LINE-like elements can induce harmful mutations (Kazazian et al. 1988; Morse et al. 1988).

At variance with the concept of "selfish DNA" is the idea that repetitive elements, while perhaps lacking inherent function, can acquire such within the genome. As a result of mutation, repetitive elements may eventually evolve coding function. Furthermore, insertion events may alter existing patterns of gene regulation in an ultimately beneficial manner (Doolittle 1989; Brosius 1991). Two cases exist in which transposable elements have been associated with mammalian gene expression: the sex-limited protein (Slp) gene of mouse is under regulatory control of a retroviral insertion acting as an androgen responsive enhancer (Stavenhagen and Robins 1988), and salivary-specific expression of human amylase may be controlled by retrotransposon insertions (Samuelson et al. (1990), and Robins and Samuelson (1992) for review).

The presence of repetitive DNA within or near coding genes has been noted in fish. *Bgl*II repeat elements exist within the spacer regions separating rDNA coding cistrons (Section 4.3). *Tc1*-like transposons are situated upstream of at least one copy

of the ependymin gene in Atlantic salmon (Section 5.2) and within introns of a vasotocin gene in hagfish (Heierhorst et al. 1992) and the IgM locus of channel catfish (Henikoff 1992). Repetitive sequences similar to LINE-like retrotransposons and to retroviral LTRs have been found upstream of histone and protamine genes in some salmonids (Moir and Dixon 1988; Winkfein et al. 1988).

In Atlantic salmon, repeat elements of different types may also exist in close association with each other. Microsatellite arrays have been found both within and near minisatellite sequences (Section 7.6 and Wright 1993). Sequence containing *NheI* repeats and features typical of DNA transposons (target site duplications and ITRs) has been inserted into the 5' flanking region of a SALT1 element (Section 6.5). Inter-SALT1 PCR suggests that Tc1-like elements are clustered at some loci (Section 5.4). Repetitive elements may promote homologous recombination and hence genomic rearrangement (see Berg and Howe (1989) for reviews), and in this manner different repeats may be brought into juxtaposition. Furthermore, a genome harbouring large numbers of inverted repeats in close proximity to each other (consider the inverted repeats within the numerous SALT1 and *NheI* elements of salmon) is probably at considerable risk of rearrangement by deletion of sequence between the repeats as a result of recombination (Gordenin et al. 1993).

What does the presence of such a genomic load of non-coding DNA with the potential for fostering genome rearrangements and altering gene expression mean for the viability of a species? Darwinian traditionalists suggest that evolution is driven by the constant, imperceptible accumulation of DNA nucleotide mutations whereby

ancestral species are slowly transformed into their descendents. Alterations in gene products may occasionally confer selective advantage upon an organism and under appropriate environmental conditions allow it to thrive. This view of the slow and steady progress of evolution came under early attack (see Huxley's letters to Darwin, Huxley 1900; Goldschmidt 1940). However, in the 1970s the challenge was elaborated into the "theory of punctuated equilibria" (Eldredge and Gould 1972) as growing fossil records and new molecular evidence exposed evolution as sometimes episodic and saltatory (Gould and Eldredge 1977). Rapid bursts of speciation may be followed by long periods of relative stasis with little evolutionary change at the species level.

Enhanced rates of mutation can arise out of genetic instability and lead to the rapid appearance of new genotypes within natural populations (see McDonald (1990) and Fontdevila (1992), and references therein). Under pressure of genomic shock (McClintock 1984) or hybridization, concerted transposition of many elements may occur. The most studied example is that of hybrid dysgenesis in *Drosophila* (Kidwell et al. 1977), whereby mobilization of P transposons is induced by the mating of males of a strain containing P elements with females lacking elements. Elevated rates of germline mutations, reversions, chromosomal aberrations and infertility are seen. In this manner, two strains may become reproductively isolated from each other. It has been suggested by Rose and Doolittle (1983) that divergent amplification of repetitive elements could also erect genetic barriers between two populations by disrupting normal chromosome pairing and recognition, thus inhibiting successful mating.

Salmonid genomes are unusually plastic and a wide range of inter-specific

Salmoninae hybrids can be produced despite chromosome numbers varying from $2n = 52$ to $2n = 84$ (Arai 1984; Dangel et al. 1973). Following a tetraploidization event (Ohno 1970), the salmonid ancestral genome underwent extensive genomic rearrangement, notably Robertsonian translocations and fusions. Against this background many repetitive DNA families have been maintained and amplified. Possible roles for repeat elements in evolution of the salmonid genome bear further investigation.

A duplicated genome, rapidly evolving by means of recombination and sequence transposition, could drive the development of new patterns of gene expression. Perhaps genome flexibility has helped the Atlantic salmon species to adapt to different ecological niches spanning a wide geographical area. The species has evolved an anadromous life style, although nonanadromous forms exist, and during its life cycle tolerates a wide range of environmental conditions of water temperature and salinity. The mechanisms by which salmon have been able to adapt and at the same time control their duplicated genome are unknown; better understanding of these processes should expand understanding of vertebrate evolution in general.

9.2. Organization of repetitive DNA families in salmonids

Sequencing of homologous repeats and zoo blot analyses provided data concerning the phylogenetic pattern of distribution of some repeat families. Certain points are worthy of note.

Interspecies variation among the Tc1-like SALT1 sequences does not indicate

a process of species-specific homogenization, as theories of molecular drive would predict. However, sampling was limited to a small number of cloned sequences from many Salmonidae species: subfamily structure may have been present but undetected. Retrotransposons, for example, typically evolve subfamilies, and these may vary in the amount by which their sequences diverge from each other (VanderWeil et al. (1993), and references therein). It is difficult to draw conclusions regarding the dynamics of repeat element evolution based on shallow sampling.

In the case of the *NheI* elements, species analysis of ASNHE and BTNHE sequences suggests the existence of subfamilies within both *Salmo* species, Atlantic salmon and brown trout (but again, more detailed sampling is required). Subfamily members are defined by nucleotide variants which they have in common. For example, the presence of subfamilies of *copia*-like retrotransposons in different species of cotton (*Gossypium* sp.) has been demonstrated (VanderWeil et al. 1993).

There exists a discontinuous phylogenetic distribution of *TcI* transposon-like sequences, which were detected in higher copy number in sturgeon and skate than in the other non-salmonid fish tested, and in *Rana* sp. to greater degree than in *Xenopus*. Several possible reasons for this phenomenon may exist:

- a) Preferential and recent amplification may have occurred in certain lineages. Inactive elements already present may have become activated.
- b) The possibility also exists that, as a result of genetic drift following speciation, transposon lineages were lost over evolutionary time. Unequal crossing-over between chromosomes might be involved in this process. *Copia*-like retrotransposons, for

example, have been severely reduced in number in *Arabidopsis* compared with other plant lineages (VanderWiel et al. 1993). Tc1 elements, while present in all strains of *C. elegans* strains, vary in copy number among strains (Emmons et al. 1983; Liao et al. 1983), and elements have been detected in *Drosophila* which also manifest species-level instability (Dowsett 1983; Hey 1989).

c) Elements may have diverged in sequence to a greater degree in some lineages than in others and hence failed to hybridize with the salmon DNA probe.

d) Horizontal transmission of elements between species may have occurred. Based on the limited sequence data available and the inability of Southern blot analysis to detect highly diverged sequence, this is not the most parsimonious explanation. The SALTI-SSal28 probe under the applied hybridization (37°C, 6x SSC (1 M [Na⁺]), 50% formamide) and wash (1x SSC, 50°C) conditions (Section 3.5) would be insensitive to nucleotide mismatch exceeding 40 percent (Sambrook et al. 1989; see also Batzer et al. 1993).

9.3. Future directions

Following upon the present description of a number of families of repetitive elements several areas for continued research are apparent:

1) As a technique for elaborating the taxonomic distribution of Tc1-like elements, PCR amplification is more sensitive than Southern blot analysis. Highly conserved sequences, detected by aligning known Tc1-like elements (Fig. 5.2, and other data noted in Chapter 5), could be used in the design of oligonucleotide primers capable

of amplifying homologous sequences from taxonomically diverse species. Cloning and sequencing of PCR products should yield information for assessing the existence and distribution of subfamilies, rates of sequence mutation, and modes of interspecies transmission.

2) As discussed in Section 1.5.2, there is evidence that the homogenization of repetitive DNA seen at the species level can also operate at the level of populations within a single species (Strachan et al. 1985; Turner et al. 1991). Therefore, satellite DNA sequence variation may reveal relationships not only between species but also between populations. However, if knowledge about the processes acting upon amplification and sequence modification of repetitive DNA is lacking, conclusions should be drawn with some caution. This caveat applies especially when the orthologous or paralogous nature of the sequences being compared is in doubt, and when there is no estimate of the frequency of multiple mutations or gene conversion. Concerted evolution may induce homoplasies, i.e. false similarities which are not ancestral. The large number of insertions and deletions common in repetitive DNA can make sequence alignment difficult.

Although more sequence data are needed, both the failure of the present study to detect species-diagnostic sequence mutations in SALT1 elements and the apparent presence of *NheI* element subfamilies both in Atlantic salmon and brown trout, suggest that sequence analysis of these repeat families may be of limited value for phylogeny reconstruction or population analysis.

On the other hand, data about the presence or absence of a transposable element

at a specific locus may be used to assess phylogenetic relationships between species (Murata et al. 1993) or, in the case of recently inserted elements, relationships between populations (Perna et al. 1992). By using oligonucleotide primers flanking a previously sequenced element, screening for insertion events by PCR is a simple matter. Sequences flanking several Atlantic salmon SALT1 elements are now available (Chapter 5, and Radice et al. 1994). Furthermore, during the course of this thesis research other SALT1 loci were isolated and stored as phage clones and these remain to be sequenced (Section 5.2).

3) Minisatellite DNA sequences isolated during this project have potential value as individual, familial, and population markers (as described in Section 6.5). A battery of probes now exists for screening by Southern blot analysis the large number of Atlantic salmon genomic DNA samples maintained in the W.S. Davidson laboratory. Useful RFLP markers may be detected. Flanking sequences for ten short sequence repeat loci are also known (Chapter 7) and may be used to design primers for PCR analysis of single locus variation.

4) The patterns of distribution of both the *Bgl*I and the *Nhe*I elements, and by extension the mechanism by which they have become established within salmonid genomes, remain to be determined (see Section 6.5). *Bgl*I repeats have been detected near rDNA genes but also exist elsewhere in the genome. Several repeat loci were isolated from the λ EMBL3 library, but only the elements contained within phage clone ASBgI22 were sequenced. Knowledge of the sequences flanking these other repeat arrays could yield information about the processes involved in their expansion.

Copies of the *NheI* element have been detected within a 1424-bp fragment inserted into a SALT1-like element. The fragment is flanked by inverted repeats and by recognition site sequences for *HpaI*. When Atlantic salmon DNA is reacted with *HpaI* it yields a UV-bright band of approximately 1470 bp that hybridizes to the *NheI* repeat (Fig. 4.1). As part of a continuing study of this element, *NheI* repeat-related DNA from this band should be cloned and sequenced.

A role for transposition in the genomic spread of some clustered tandem repeats has been proposed for a few organisms. For example, transposition insertion events appear to have been involved in the widespread dispersion of *Cla* elements (120-bp tandem repeat sequences) within the genome of a midge subspecies (Hankeln et al. 1994). *Cla* repeats are present in *Chironomus thummi thummi* but absent in related subspecies at some orthologous loci (including the non-transcribed spacer region of an rDNA gene). The two possible modes of mobilization of *Cla* elements proposed by Hankeln et al. (1994) may also be relevant for the salmon *BglI* or *NheI* elements: retroposition involving RNA intermediates, or transposition by excision and reintegration. As noted in Sections 1.3.1 and 1.3.2.3, both RNA transcripts and extrachromosomal copies of satellite-like DNA sequence have been detected in various organisms.

5) The long-range genomic organization of repeat loci, including their chromosomal distribution, may be investigated by *in situ* chromosome hybridization and pulsed field gel electrophoresis. Different families of satellite DNA may be detected by comparing periodicity of large restriction enzyme fragments on Southern blots

following PFGE (Wevrick and Willard 1989; Marçais et al. 1991, Modi et al. 1993).

Pendás et al. (1994c) demonstrated that *Bgl*II repeat sequence can identify the NOR-bearing pair of chromosomes in Atlantic salmon by fluorescent *in situ* hybridization. Other repeat types discussed above may be found to be limited in their genomic distribution, and when used as probes in FISH studies may ultimately serve as valuable chromosome-specific markers for cytogenetic analyses.

6) Repetitive DNA elements could potentially be developed as transgenic vectors. He et al. (1992) linked tandem repeats from zebrafish to an *E. coli* chloramphenicol acetyltransferase (CAT) reporter gene and microinjected copies of the plasmid construct into oocytes. Evidence for integration was ambiguous, but presumably any such event would occur as a result of recombination of the construct with pre-existing repeat copies.

Use of SALT1 sequences as the basis for an interspecies genetic transformation system may hold promise. Gloor et al. (1991) have demonstrated the feasibility of *Drosophila* P-element induced gene replacement, and suggest that other class II transposons which generate double-stranded gaps during transposition, including Tc1, might also prove amenable to this technique. However, the development of Tc1-like elements as transgene vectors for fish and perhaps other vertebrates may be predicated on isolating an intact element copy, one lacking frameshifts, deletions, and stop codons. The fact that all Tc1-like elements so far obtained from fish have been defective, suggests that this may prove a difficult task. However, defective elements could be "repaired", perhaps by using techniques of site-directed mutagenesis. A recent review

by Lai (1994) outlines potentially exciting applications of transposons in genetic research.

REFERENCES

- Allendorf, F.W., and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes. *In* Evolutionary Genetics of Fishes. (Turner, B.J., ed.). Plenum Press, New York. pp 1-53.
- Altshul, S.F., Gish, W., Miller, W., Myers, M.W., and Lipman, D.J. 1990. A basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Amemiya, C.T., and Gold, J.R. 1986. Chromomycin A₃ stains nucleolus organizer regions of fish chromosomes. *Copeia* 1: 226-231.
- Amos, W., and Dover, G.A. 1991. The use of satellite DNA sequences in determining population differentiation in the minke whale. *Reports of the International Whaling Commission, Special Issue* 13: 99-104.
- Arai, K. 1984. Developmental genetic studies on salmonids: morphogenesis, isozyme phenotypes and chromosomes in hybrid embryos. *Mem. Fac. Fish., Hokkaido Univ.* 31: 1-94.
- Armour, J.A.L., and Jeffreys, A. 1992. Recent advances in minisatellite biology. *FEBS Letts.* 307: 113-115.
- Armour, J.A.L., Crosier, M., and Jeffreys, A.L. 1992. Human minisatellite alleles detectable only after PCR amplification. *Genomics* 12: 116-124.
- Armour, J.A.L., Wong, Z., Wilson, V., Royle, N.J., and Jeffreys, A.J. 1989. Sequences flanking repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. *Nucl. Acids Res.* 17: 4925-4935.
- Arnheim, N. 1983. Concerted evolution of multigene families. *In* Evolution of Gene and Proteins. (Nei, M. and Koehn, K., eds.). Sinauer, Sunderland, MA. pp. 38-61.
- Batzer, M.A., Schmid, C.W., and Deininger, P.L. 1993. Evolutionary analyses of repetitive DNA sequences. *Meth. Enzym.* 224: 213-232.
- Beckmann, J.S., and Weber, J.L. 1992. Survey of human and rat microsatellites. *Genomics* 12: 627-631.
- Behnke, R.J. 1992. Naïve trout of western North America. *Am. Fish. Soc. Mono.* 6.
- Bentzen, P., and Wright, J.M. 1993. Nucleotide sequence and evolutionary conservation of a minisatellite variable number tandem repeat cloned from Atlantic

salmon, *Salmo salar*. Genome 36: 271-277.

Bentzen, P., Harris, A.S., and Wright, J.M. 1991. Cloning of hypervariable minisatellite and simple sequence microsatellite repeats for DNA fingerprinting of important aquacultural species of salmonids and tilapia. In DNA Fingerprinting: Approaches and Applications. (Burke, T., Dolf, G., Jeffreys, A.J., and Wolff, eds.). Birkhauser Verlag, Basel, Switzerland. pp. 243-262.

Bentzen, P., Taylor, E.B., and Wright, J.M. 1993. A novel synthetic probe for DNA fingerprinting salmonid fishes. J. Fish Biol. 43: 313-316.

Berg, D.E., and Howe, M.M. (eds.). 1989. Mobile DNA, American Society for Microbiology, Washington, D.C.

Bermingham, E., Forbes, S.H., Friedland, K., and Pla, C. 1991. Discrimination between Atlantic salmon (*Salmo salar*) of North American and European origin using restriction analyses of mitochondrial DNA. Can. J. Fish. Aquat. Sci. 48: 884-893.

Bernardi, G., and Bernardi, G. 1990. Compositional patterns in the nuclear genome of cold-blooded vertebrates. J. Mol. Evol. 31: 265-281.

Birt, T.M., Green, J.M., and Davidson, W.S. 1991. Mitochondrial DNA variation reveals genetically distinct sympatric populations of anadromous and nonanadromous Atlantic salmon, *Salmo salar*. Can. J. Fish. Aquat. Sci. 48: 577-582.

Blanchetot, A. 1991. A *Musca domestica* satellite sequence detects individual polymorphic regions in insect genome. Nucl. Acids Res. 19: 929-932.

Blaxhall, P.C. 1983. Chromosome karyotyping of fish using conventional and G-banding methods. J. Fish. Biol. 22: 417-424.

Borkhsenius, S.N., and Chernov, V.M. 1988. *Sau3A* repeat sequences in genomes of two forms of sockeye salmon. Mol. Biol. (Moscow) 22: 439-445.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction length polymorphisms. Am. J. Hum. Genet. 32: 314-31.

Boyle, A.L., Feltquite, D.M., Dracopoli, N.C., Housman, D.E., and Ward, D.C. 1992. Rapid physical mapping of cloned DNA on banded mouse chromosomes by fluorescence *in situ* hybridization. Genomics 12: 106-115.

Brenig, B., and Brem, G. 1991. Human VNTR sequences in porcine HTF-islands.

Experientia 58: 39-49.

Brenner, S., Elgar, G., Sandford, R., Macrea, A., Venkatesh, B., and Aparicio, S. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* 266: 265-268.

Brezinsky, L., Wang, G.V., Humphreys, T., and Hunt, J.F. 1990. The transposable element Uhu from Hawaiian *Drosophila* - member of the widely dispersed class of Tc1-like transposons. *Nucl. Acids Res.* 18: 2053-2059.

Britten, R.J., and Kohne, D.E. 1968. Repeated sequences in DNA. *Science* 161: 529-540.

Brooker, A.L., Cook, D., Bentzen, P., Wright, J.M., and Doyle, R.W. 1994. The organization of microsatellites differs between mammals and cold-water teleost fishes. *Can. J. Fish. Aquat. Sci.* (in press).

Brosius, J. 1991. Retrotransposons - seeds of evolution. *Science* 251: 753.

Broun, P., and Tanksley, S.D. 1993. Characterization of tomato DNA clones with sequence similarity to human minisatellites 33.6 and 33.15. *Plant Molec. Biol.* 23: 231-242.

Buchanan, F.C., Adams, L.J., Littlejohn, R.P., Maddox, J.F., and Crawford, A.M. 1994. Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* 22: 397-403.

Cabot, E.L., and Beckenbach, A.T. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comput. Appl. Biosci.* 5: 233-234.

Caizzi, R., Caggese, C., and Pimpinelli, S. 1993. Bari-1, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. *Genetics*. 122: 335-345.

Cantor, C.R., Smith, C.L., and Mathew, M.K. 1988. Pulsed-field gel electrophoresis of very large DNA molecules. *Ann. Rev. Biophys. Biophys. Chem.* 17: 287-304.

Carver, R.E., Mair, G.C., Skibinski, D.O.F., Parkin, D.T., and Beardmore, J.A. 1991. The application of DNA fingerprinting in the analysis of gynogenesis of tilapia. *Aquaculture* 95: 41-52.

Cau, A., Salvadori, S., Deiana, A.M., Bella, J.L., and Mezzanotte, R. 1988. The characterization of *Muraena helena* L. mitotic chromosomes: karyotype, C-banding,

nucleolar organizer regions, and *in situ* digestion with restriction endonucleases. Cytogenet. Cell Genet. 47: 223-226.

Choo, K.H., Vissel, B., Nagy, A., Earle, E., and Kalitsis, P. 1991. A survey of genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. 19: 1179-1182.

Clark, J.B., Maddison, W.P., and Kidwell, M.G. 1994. Phylogenetic analysis supports horizontal transfer of P transposable elements. Mol. Biol. Evol. 11: 40-50.

Collins, J. 1980. The instability of palindromic DNA. Cold Spring Harbor Symp. Quant. Biol. 45: 409-416.

Collins, J.J., and Anderson, P. 1994. The Tc5 family of transposable elements in *Cuenorhabditis elegans*. Genetics 137: 771-781.

Coppieters, W., Van de Weghe, A., Depicker, A., Bouquet, Y., and VanEveren, A. 1990. A hypervariable pig DNA fragment. Animal Genet. 21: 29-38.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16: 10881-10890.

Cox, R.D., Copeland, N.G., Jenkins, N.A., and Lehrach, H. 1991. Interspersed repetitive element polymerase chain reaction product mapping using a mouse interspecific backcross. Genomics 10: 375-84.

Cox, R.D., and Lehrach, H. 1991. Genome mapping: PCR based meiotic and somatic cell hybrid analysis. BioEssays 13: 193-97.

Cutler, M.G., Bartlett, S.E., Hartley, S.E., and Davidson, W.S. 1991. A polymorphism in the ribosomal RNA genes distinguishes Atlantic salmon (*Salmo salar*) from North America and Europe. Can. J. Fish. Aquat. Sci. 48: 1655-1661.

Dangel, J.R., Macy, P.T., and Withler, F.C. 1973. Annotated bibliography of interspecific hybridization of fishes of the subfamily Salmoninae. NOAA Tech. Memo. NMFS NWFC-1, Seattle.

Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A. 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. Genetics. 124: 339-355.

Das, H.K., Jackson, C.L., Miller, D.A., Leff, T., and Breslow, J.L. 1987. The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific

minisatellite in its third intron. *J. Biol. Chem.* 262: 4787-4793.

Datta, U., Dutta, P., and Mandel, R.K. 1988. Cloning and characterization of a highly repetitive fish nucleotide sequence. *Gene* 62: 331-336.

Davidson, W.S., Birt, T.P., and Green, J.M. 1989. A review of genetic variation in Atlantic salmon, *Salmo salar* L., and its importance for stock identification, enhancement programmes and aquaculture. *J. Fish Biol.* 34: 547-560.

Davis, R.W., Botstein, D., and Roth, J.R. 1980. *A Manual of Genetic Engineering: Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Dawid, I.B., D.D. Brown, and R.H. Reeder. 1970. Composition and structure of chromosomal and amplified ribosomal DNA's of *Xenopus laevis*. *J. Mol. Biol.* 51: 341-360.

Deininger, P.L. 1989. SINEs: short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA*. (Berg, D.E. and Howe, M.M., eds.). American Society for Microbiology, Washington, D.C. pp. 537-556.

Delany, M.E., and Bloom, S.E. 1984. Replication banding patterns in the chromosomes of the rainbow trout. *J. Hered.* 75: 431-434.

Denovan, E.M., and Wright J.M. 1990. A satellite DNA family from pollock (*Pollachius virens*). *Gene* 87: 279-283.

Devlin, R.H., McNeil, B.K., Groves, T.D.D., and Donaldson, E.M. 1991. Isolation of a Y-chromosomal DNA probe capable of determining genetic sex in chinook salmon (*Oncorhynchus tshawytscha*). *Can. J. Fish. Aquat. Sci.* 48: 1606-1612.

Doak, T.G., Doerder, F.P., Jahn, C.L., and Herrick, G. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D25E" motif. *Proc. Natl. Acad. Sci. USA.* 91: 942-946.

Doolittle, W.F. 1989. Heirarchical approaches to genome evolution. *Can. J. Phil. Suppl.* 14: 101-133.

Doolittle, W.F., and Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.

Doolittle, R.F., Feng, D.-F., Johnson, M.S., and McClure, M.A. 1989. Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol.* 64: 1-30.

- Dover, G.A. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.
- Dowsett, A.P. 1983. Closely related species of *Drosophila* can contain different libraries of middle repetitive sequences. *Chromosoma* 88: 104-108.
- Egger, K.N., and Sigler, S. 1993. Relatedness of the ericoid endophyte *Scytalidium vaccinii* and *Hymenoscyphus ericae* inferred from the analysis of ribosomal DNA. *Mycologia* 85: 219-230.
- Ekker, M., Fritz, A., and Westerfield, M. 1992. Identification of two families of satellite-like repetitive DNA sequences from the zebrafish (*Brachydanio rerio*). *Genomics* 13: 1169-1173.
- Elder, J.F., Jr., and Turner, B.J. 1994. Concerted evolution at the population level: pupfish *HindIII* satellite DNA sequences. *Proc. Natl. Acad. Sci. USA*. 91: 994-998.
- Eldredge, N., and Gould, S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In *Models in Paleobiology* (Schopf, T.J.M., ed.). Freeman, Cooper, San Francisco. pp. 82-115.
- Emmons, S.W., Yesner, L., Ruan, K.S., and Katzenberg, D. 1983. Evidence for a transposon in *Caenorhabditis elegans*. *Cell* 32: 55-65.
- Engels, W.R., Johnson-Schlitz, D.M., Eggeleston, W.B., and Sved, J. 1989. High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* 62: 515-525.
- Epstein, L.M., Mahom, K.A., and Gall, J.G. 1986. Transcription of a satellite DNA in the newt. *J. Cell Biol.* 103: 1137-44.
- Estoup, A., Presa, P., Krieg, F., Vaiman, D., and Guyomard, R. 1993. (CT)_n and (GT)_n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* 71: 488-498.
- Felsenstein, J. 1991. PHYLIP manual, version 3.4. University of Washington, Seattle.
- Finnegan, D.J. 1985. Transposable elements in eukaryotes. *Int. Rev. Cytol.* 93: 281-326.
- Finnegan, D.J. 1989. Eukaryotic transposable elements and genome evolution. *TIG* 5: 103-107.
- Flavell, A.J., and Smith, D.B. 1992. A Ty-1-copia group retrotransposon sequence in

a vertebrate. Mol. Gen. Genet. 233: 322-326.

Fontdevila, A. 1992. Genetic instability and rapid speciation: are they coupled? Genetica 86: 247-258.

Franck, J.P.C. 1993. Organization and evolution of two satellite DNA families, SATA and SATB, from the tilapiine and haplochromine genome (Pisces: Cichlidae). PhD thesis, Dept. Biology, Dalhousie Univ. 143 pp.

Franck, J.P.C., and Wright, J.M. 1993. Conservation of satellite DNA sequence (SATB) in the tilapiine and haplochromine genome (Pisces: Cichlidae). Genome 36: 187-194.

Franck, J.P.C., Wright, J.M., and McAndrew, B.J. 1992. Genetic variability in a family of satellite DNAs from tilapia (Pisces: Cichlidae). Genome 35: 719-725.

Franck, J.P.C., Kornfield, I., and Wright, J.M. 1994. The utility of SATA satellite DNA sequences for inferring phylogenetic relationships among the three major genera of tilapiine cichlid fishes. Mol. Phylogenet. Evol. 3: 10-16.

Franz, G., and Savakis, C. 1991. *Minos*, a new transposable element from *Drosophila hydei* is a member of the Tc1-like family of transposons. Nucl. Acids Res. 19: 6646.

Fujii, H., and Y. Ojima. 1983. Chromosome Data Retrieval System CDR: end-user language and its application to fish cytotaxonomy. Kwansei Gakuin Univ. Ann. Stud. 32. (in Japanese).

Garrido-Ramos, M.A., Jamilena, M., Lozano, R., Ruiz Rejón, C., and Ruiz Rejón, M. 1994. Cloning and characterization of a fish centromeric satellite DNA. Cytogenet. Cell. Genet. 65: 233-237.

Gharret, A.J., Simon, R.C., and McIntyre, J.D. 1977. Reassociation and hybridization properties of DNAs from several species of fish. Comp. Bioch. Physiol. 56B: 81-85.

Giles, V., Thode, G., and Alvarez, M.C. 1988. Early replication bands in two scorpion fishes, *Scorpaena porcus* and *S. notata* (order Scorpaeniformes). Cytogenet. Cell Genet. 47: 80-83.

Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R., and Engels, W.R. 1991. Targeted gene replacement in *Drosophila* via P element-induced gap repair. Science 253: 1110-1117.

Goff, D.J., Galvin, K., Katz, H., Westerfield, M., Lander, E., and Tabin, C.J. 1992.

Identification of polymorphic simple sequence repeats in the genome of the zebrafish. *Genomics* 14: 200-202.

Goldschmidt, R.B. 1940. *The Material Basis of Evolution*. Yale Univ. Press, New Haven. 436 pp.

Gonnet, G.H., Cohen, M.A., and Benner, S.A. 1992. Exhaustive matching of the entire protein sequence database. *Science* 256: 1443-1445.

Goodier, J.L., and Davidson, W.S. 1993a. Gene mapping in fish. *In Biochemistry and Molecular Biology of Fishes*. Vol. 2. (Hochachka, P., and Mommsen, T., eds.). Elsevier, New York. pp. 93-110.

Goodier, J.L., and Davidson, W.S. 1993b. A repetitive element in the genome of Atlantic salmon, *Salmo salar*. *Gene* 121: 237-242.

Goodier, J.L., and Davidson, W.S. 1994a. Tc1 transposon-like sequences are widely distributed in salmonids. *J. Mol. Biol.* 241: 26-34.

Goodier, J.L., and Davidson, W.S. 1994b. Characterization of a repetitive element detected by *NheI* in the genomes of *Salmo* species. *Genome* 37: 639-645.

Goodier, J.L., and Davidson, W.S. 1994c. Characterization of families of minisatellite repeats in Atlantic salmon (*Salmo salar*). *Molecular Biology and Evolution* (submitted).

Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E., and Resnick, M.A. 1993. Inverted DNA repeats: a source of eukaryotic genomic instability. *Molec. Cell. Biol.* 13: 5315-5322.

Gould, S.J., and Eldredge, N. 1977. Punctuated equilibria: the tempo and mode of evolution revisited. *Paleobiology* 3: 115-151.

Grewe, P.D., Billington, N., and Hebert, P.D.N. 1990. Phylogenetic relationships among members of *Salvelinus* inferred from mitochondrial DNA divergence. *Can. J. Fish. Aqu. Sci.* 47: 984-991.

Gross, M.L., Kapuscinski, A.R., and Faras, A.J. 1994. Nest-specific DNA fingerprints of smallmouth bass in Lake Opeongo, Ontario. *Trans. Am. Fish. Soc.* 123: 449-459.

Gyllenstein, U. 1985. The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous, and freshwater species. *J. Fish. Biol.* 26: 691-699.

- Gyllenstein, U.B., Jakobsson, S., Temrin, H., and Wilson, A.C. 1989. Nucleotide sequence and genomic organization of bird minisatellites. *Nucl. Acids Res.* 17: 2203-2214.
- Hallerman, E.M., and Beckmann, J.S. 1988. DNA-level polymorphism as a tool in fisheries science. *Can. J. Fish. Aquat. Sci.* 45: 1075-1087.
- Hamada, H., Petrino, M.G., Kakunaga, T., Seidman, M., and Stollar, B.D. 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA.* 79: 6465-69.
- Hanham, A.F. and Smith, M.J. 1980. Sequence homology in the single-copy DNA of salmon. *Comp. Biochem. Physiol.* 65B: 333-338.
- Hankeln, T., Rohwedder, A., Weich, B., and Schmidt, E.R. 1994. Transposition of minisatellite-like DNA in *Chironomus* midges. *Genome* 37: 542-549.
- Harris, A.S., Bieger, S., Doyle, R.W., and Wright, J.M. 1990. DNA fingerprinting of tilapia, *Oreochromis niloticus*, and its application to aquaculture genetics. *Aquaculture* 92: 157-163.
- Harris, L.J., Prasad, S.S., and Rose, A.M. 1990. Isolation and sequence analysis of *C. brigssae* repetitive elements related to the *C. elegans* transposon Tc1. *J. Mol. Evol.* 30: 359-369.
- Hartley, S.E. 1987. The chromosomes of salmonid fishes. *Biol. Rev.* 62: 197-214.
- Hartley, S.E. 1991. Restriction enzyme banding in Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*). *Genet. Res.* 57: 273-278.
- Hartley, S.E., and Horne, M.T. 1984. Chromosome relationships in the genus *Salmo*. *Chromosoma* 90: 229-237.
- Hartley, S.E., and Davidson, W.S. 1994. Characterization and distribution of genomic repeat sequences from Arctic charr (*Salvelinus alpinus*). In *Genetics and Evolution of Aquatic Organisms*. (Beaumont, A.R., ed.). Chapman and Hall, London. pp. 271-280.
- Hauser, L., Carvalho, G.R., Hughes, R.N., and Carter, R.E. 1992. Clonal structure of the introduced freshwater snail *Posamopyrgus antipodarum* (Prosobranchia: Hydrobiidae), as revealed by DNA fingerprinting. *Proc. Roy. Soc. Lond., B* 249: 19-25.
- Heath, D.D., Iwama, G.K., and Devlin, R.H. 1993. PCR primed with VNTR core

sequences yields species specific patterns and hypervariable probes. Nucl. Acids Res. 21: 5782-5785.

Hedrick, R.P., and Wingfield, W.H. 1991. A small RNA virus isolated from salmonid fishes in California, U.S.A. Can. J. Fish Aquat. Sci. 48: 99-104.

Heierhorst, J., Lederis, K. and Richter, D. 1992. Presence of a member of the Tc1-like transposon family from nematodes and *Drosophila* within the vasotocin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*. Proc. Natl. Acad. Sci. USA. 89: 6798-6802.

He, L., Li, G., Xie, Y., and Zhu, Z. 1987. Common carp repetitive sequence cloning and characterization. Res. Coll., Hubei Prov. Biochem. Soc.: 18-23.

He, L., Zhu, Z., Faras, A.J., Guise, K.S., Hackett, P.B., and Kapuscinski, A.R. 1992. Characterization of *AluI* repeats of zebrafish (*Brachydanio rerio*). Mol. Marine Biol. Biotech. 1: 125-135.

Hellmer, A., Voiculescu, I., and Schempp, W. 1991. Replication banding studies in two cyprinid fishes. Chromosoma 100: 524-531.

Henikoff, S. 1992. Detection of *Caenorhabditis* transposon homologs in diverse organisms. New Biologist 4: 382-388.

Hey, J. 1989. The transposable portion of the genome of *Drosophila algonguin* is very different from that in *D. melanogaster*. Mol. Biol. Evol. 8: 282-296.

Hinegardner, R., and Rosen, D.E. 1972. Cellular DNA content in the evolution of teleostan fishes. Am. Nat. 106: 621-644.

Holmquist, G.P. 1989. Evolution of chromosome bands: molecular ecology of noncoding DNA. J. Mol. Evol. 28: 469-486.

Houck, M.A., Clark, J.B., Peterson, K.R., and Kidwell, M.G. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science 253: 1125-1129.

Huang, C.-J., Huang, F.-L., Wang, Y.-C., Hsiao, P.-W., and Lo, T.-B. 1993. Molecular cloning and relationship of highly repetitive *HindIII* sequences in three cyprinid species: silver carp, bighead carp and grass carp. Proc. Natl. Sci. Council, ROC (Pt. B) 17: 85-90.

Huxley, T.H. 1900. Life and Letters of Thomas Henry Huxley, by his son Leonard

Huxley. 2 vols. London: Macmillan.

Hyrien, O., Debatisse, M., Buttin, G., and de Saint Vincent, B.R. 1987. A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A+T-rich DNA. *EMBO J.* 6: 2401-2408.

Inglehearn, C.F., and Cooke, H.J. 1990. A VNTR immediately adjacent to the human pseudoautosomal telomere. *Nucl. Acids Res.* 18: 471-476.

Iwasaki, H., Stewart, P.W., Dilley, W.G., Holt, M.S., Steinbrueck, T.D., Wells, S.A. Jr., and Donis-Keller, H. 1992. A minisatellite and a microsatellite polymorphism within 1.5 kb at the human muscle glycogen phosphorylase (PYGM) locus can be amplified by PCR and have combined informativeness of PIC 0.95. *Genomics* 13: 7-15.

Jacobson, J.W., Guo, W., and Hughes, C.R. 1992. A *Drosophila* minisatellite contains multiple *chi* sequences. *Insect Biochem. Molec. Biol.* 22: 785-792.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985a. Hypervariable "minisatellite" regions in human DNA. *Nature* 314: 67-73.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985b. Individual-specific "fingerprints" of human DNA. *Nature* 316: 76-79.

Jones, R.S., and Potter, S.S. 1985. Characterization of cloned human alphoid satellite with an unusual monomeric polydispersed circular DNA. *Nucl. Acids Res.* 13: 1027-42.

Jorgenson, A.L., C.J. Bostock, and A.L. Bak. 1986. Chromosome-specific subfamilies within human alphoid repetitive DNA. *J. Mol. Biol.* 187: 185-96.

Kaplan, F.S., Murray, J., Sylvester, J.E., Gonzalez, I.L., O'Connor, J.P., Doering, J.L., Muenke, M., Emanuel, B.S., and Zasloff, M.A. 1993. The topographic organization of repetitive DNA in the human nucleolus. *Genomics* 15: 123-132.

Karess, R.E., and Rubin, G.M. 1984. Analysis of P transposable element functions in *Drosophila*. *Cell* 38: 135-146.

Kashi, Y., Iraqi, F., Tikochinski, Y., Ruzitsky, B., Nave, A., Beckmann, J.S., Friedmann, A., Soller, M., and Gruenbaum, Y. 1990. (TG)_n uncovers a sex-specific hybridization pattern in cattle. *Genomics* 7: 31-36.

Kass, D.H., Batzer, M.A., and Deininger, P.L. 1994. Gene conversion as a secondary mechanism of SINE evolution. *Mol. Cell Biol.* (in press).

- Kazazian, H.H.J., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., Antonarakis, S.E. 1988. Haemophilia A resulting from *de novo* insertion of LI sequences represents a novel mechanism for mutation in man. *Nature* 332: 164-166.
- Kelly, R., Bulfield, G., Collick, A., Gibbs, M., and Jeffreys, A.J. 1989. Characterization of a highly unstable mouse minisatellite locus: evidence for somatic mutation during early development. *Genomics* 5: 844-856.
- Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M. and Okada, N. 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retrotransposons during evolution. *Proc. Natl. Acad. Sci. USA*. 88: 2326-2330.
- Kidwell, M.G., Kidwell, J.F., and Sved, J.A. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics* 86: 813-833.
- Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 11-120.
- King, C.C. 1992. Modular transposition and the dynamical structure of eukaryote regulatory evolution. *Genetica* 86: 127-142.
- Kit, S. 1961. Equilibrium centrifugation in density gradients of DNA preparations from animal tissues. *J. Mol. Biol.* 3: 711-716.
- Kitazawa, T., Kominami, R., Tanaka, R., Wakabayashi, K., Nagao, M., Signer, E.N., Gu, F., Gustavsson, I., Andersson, L., and Jeffreys, A.J. 1994. A pseudoautosomal minisatellite in the pig. *Mammalian Genome* 5: 48-51.
- Kiyama, R., Okumura, K., Matsui, H., Bruns, G.A.P., Kanda, N., and Oishi, M. 1987. Nature of recombination involved in excision and rearrangement of human repetitive DNA. *J. Mol. Biol.* 198: 589-598.
- Kominami, R., Mitani, K., and Muranatsu, K. 1988. Nucleotide sequence of a mouse minisatellite DNA. *Nucl. Acids. Res.* 16: 1197.
- Kubota, S., Kuro-o, M., Mizuno, S., and Kohno, S. 1993. Germ line-restricted, highly repeated DNA sequences and their chromosomal localization in a Japanese hagfish (*Eptatretus okinoseanus*). *Chromosoma* 102: 163-173.
- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G. and Skalka, A.M. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved

- among retroviral/introtransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12: 2332-2338.
- Lai, C. 1994. Genetic applications of transposable elements in eukaryotes. *Genome* 37: 519-525.
- Lambowitz, A.M. 1989. Infectious introns. *Cell* 56: 323-326.
- Lande, R., and Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743-756.
- Laughlin, T.F., and Turner, B.J. 1994. Multilocus DNA fingerprinting detects population differentiation in the outbred and abundant fish species *Poecilia latipinna*. *Mol. Ecol.* 3: 263-266.
- Laursen, H.B., Jørgensen, A.L., Jones, C., and Bak, A.L. 1992. Higher rate of X chromosome α -repeat DNA in human than in the great apes. *EMBO J.* 11: 2367-2372.
- Lawrence, J.B., Singer, R.H., and McNeil, J.A. 1990. Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* 249: 928-932.
- Ledbetter, S.A., Nelson, D.L., Warren, S.T., and Ledbetter, D.H. 1990. Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* 6: 475-481.
- Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203-21.
- Levinson, G., Marsh, J.L., Epley, J.T., and Gutman, G.A. 1985. Cross-hybridizing snake satellite, *Drosophila* and mouse sequences may have arisen independently. *Mol. Biol. Evol.* 2: 494-504.
- Liao, L., Rosenzweig, B., and Hirsch, D. 1983. Analysis of a transposable element in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA.* 80: 3585-3589.
- Liu, L. 1986. Studies on making fish chromosomes elongated with high resolution G-banding. *In* Indo-Pacific Fish Biology: Proceedings of the Second International Conference on Indo-Pacific Fishes. (Ueno, T., Arai, R., Taniuchi, T., and Matsunura, K. eds.). Ichthyological Society of Japan. pp. 910-917.
- Lloyd, M.A., and Fields, M.J., and Thorgaard, G.H. 1989. Bkm minisatellites are not sex associated but reveal DNA fingerprint polymorphisms in rainbow trout. *Genome*

Lloyd, M.A., and Thorgaard, G.H. 1988. Restriction endonuclease banding of rainbow trout chromosomes. *Chromosoma* 96: 171-177.

Love, J.M., Knight, A.M., McAleer, M.A., and J.A. Todd. 1990. Towards construction of the high resolution map of the mouse genome using PCR-analyzed microsatellites. *Nucl. Acids Res.* 18: 4123-4130.

Lozano, R., Ruiz Rejón, C., and Ruiz Rejón, M. 1991. An analysis of coho salmon chromatin by means of C-banding, AG- and fluorochrome staining, and *in situ* digestion with restriction endonucleases. *Heredity* 66: 403-409.

Lynch, M. 1988. Estimation of relatedness by DNA fingerprinting. *Mol. Biol. Evol.* 5: 584-599.

Maio, J.J. 1976. Properties of selected eukaryotic satellite DNAs. *In Handbook of Biochemistry, Nucleic Acids*, v. 2. (Freeman, G.D., ed.). Chemical Rubber Co., Cleveland. pp. 379-404.

Manuelidis, L. 1978. Complex and simple sequences in human repeated DNAs. *Chromosoma* 66: 1-21.

Marçais, B., Laurent, A.-M., Charlieu, J.-P., and Koizès, G. 1993. Organization of the variant domains of α satellite DNA on human chromosome 21. *J. Mol. Evol.* 37: 171-178.

Martienssen, R.A., and Baulcombe, D.C. 1989. An unusual wheat sequence (WIS1) lies upstream of an α -amylase gene in hexaploid wheat, and carries a "minisatellite" array. *Mol. Gen. Genet.* 217: 401-410.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* 109: 1963-1973.

May, B., and Johnson, K.R. 1990. Composite linkage map of salmonid fishes (*Salvelinus*, *Salmo*, *Oncorhynchus*) *In Genetic Maps: Locus Maps of Complex Genomes*. Book 4. Nonhuman Vertebrates. (S.J. O'Brien, ed.). Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y. pp. 151-159.

Mayr, B., Kalat, M., and Rab, P. 1986. Localization of NORs and counterstain-enhanced fluorescence studies in *Salmo gairdneri* and *Salmo trutta*. *Theor. Appl. Genet.* 71: 703-707.

- McClintock, B. 1984. The significance of the responses of the genome to challenge. *Science* 226: 792-801.
- McDonald, J.F. 1990. Macroevolution and retroviral elements. *BioScience* 40: 183-191.
- McGowan, C., and Davidson, W.S. 1992. Artificial hybridization of Newfoundland brown trout and Atlantic salmon: hatchability, survival and growth to first feeding. *Aquaculture* 106: 117-125.
- McKusick, V.A. 1991. Current trends in mapping human genes. *FASEB J.* 5: 12-20.
- McVeigh, H.P., Bartlett, S.E., and Davidson, W.S. 1991. Polymerase chain reaction/direct sequence analysis of the cytochrome b gene in *Salmo* species. *Aquaculture* 95: 225-233.
- Medrano, L., Bernardi, G., Couturier, J., Dutrillaux, B., and Bernardi, G. 1988. Chromosome banding and genome compartmentalization in fishes. *Chromosoma* 96: 178-183.
- Miklos, G.L.G., and Gill, A.C. 1982. Nucleotide sequences of highly repeated DNAs: compilation and comments. *Genet. Res.* 39: 1-30.
- Miller, S.A., Dykes, D.D. and Polesky, H.F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl. Acids Res.* 16: 1215.
- Mitchell, A.R., Gosden, J.R., and Ryder, O.A. 1981. Satellite DNA relationships in man and primates. *Nucl. Acids. Res.* 9: 3235-3249.
- Modi, W.S. 1993. Heterogeneity in the concerted evolution process of a tandem satellite array in meadow mice (*Microtus*). *J. Mol. Evol.* 37: 48-56.
- Moerman, D.G. and Waterston, R.H. 1989. Mobile elements in *Caenorhabditis elegans* and other nematodes. In *Mobile DNA* (Berg, D.E. and Howe, M.M., eds). American Society for Microbiology, Washington, D.C. pp. 537-556.
- Moir, R.D. 1988. Structure of several multigene families in salmonid families. PhD thesis. Dept. Med. Biochem., Univ. Calgary. 263 pp.
- Moir, R.D. and Dixon, G.H. 1988. A repetitive DNA sequence in the Salmonid fishes similar to a retroviral long terminal repeat. *J. Mol. Evol.* 27: 1-7.
- Morse, B., Rotherg, P.G., South, V.J., Spandorfer, J.M., and Astrin, S.M. 1988.

Insertional mutagenesis of the *myc* locus by a LINE-I sequence in human breast carcinoma. *Nature* 333: 87-90.

Moyer, S.P., Ma, D.P., Thomas, T.L., and Gold, J.R. 1988. Characterization of a highly repeated satellite DNA from the cyprinid fish *Notropis lutrensis*. *Comp. Biochem. Physiol.* 91B: 639-646.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.-R. 1988. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA*. 85: 6622-6626.

Müller-Schmid, A., Rinder, H., Lottspeich, F., Gertzen, E.-M., and Hoffman, W. 1992. Ependymins from the cerebrospinal fluid of salmonid fish: gene structure and molecular characterization. *Gene*. 118: 189-196.

Murata, S., Takasaki, N., Saitoh, M., and Okada, N. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short, interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA*. 90: 6995-6999.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-1622.

Nanda, I., Feichtinger, W., Schmid, M., Schroder, J.J., Zischler, H., and Epplen, J.T. 1990. Simple repetitive sequences are associated with differentiation of the sex chromosomes in the guppy fish. *J. Mol. Evol.* 30: 456-462.

Nei, M., and Tajima, F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H., and Caskey, T. 1989. *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA*. 86: 6686-6690.

Nelson, J.S. 1994. *Fishes of the World*. 3rd ed. John Wiley and Sons, Ltd., New York.

Neuer-Nitsche, B., Lu, X., and Werner, D. 1988. Functional role of a highly repetitive DNA sequence in anchorage of the mouse genome. *Nucl. Acids Res.* 16: 8351-60.

- Okamura, K., Kiyama, R., and Oishi, M. 1987. Sequence analyses of extrachromosomal *Sau3A* and related family DNA: analysis of recombination in the excision event. *Nucl. Acids Res.* 15: 7477-7489.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, N.Y.
- Ohno, S. 1969. So much "junk" data in our genome. *In* *Evolution of Genetic Systems*. (Smith, H.H., ed.). Brookhaven Symp. 26. Gordon-Breach, New York. pp. 366-370.
- Ohno, S., Muramoto, J., Klein, J. and Atkin, N.B. 1969. Diploid-tetraploid relationship in clupeoid and salmonid fish. *In* *Chromosomes Today*, V. II (Darlington, C.P., and Lewis, K.R., eds.). Oliver and Boyd, Edinburgh. pp. 139-147.
- Ojima, Y. 1983. *Fish Cytogenetics*. Kwansei Gakuin Univ. Ann. Stud. 47. 247 pp. (in Japanese).
- Orgel, L.E., and Crick, F.H.C. 1980. Selfish DNA: the ultimate parasite. *Nature* 284: 604-607.
- Padilla, J.A., Fernández-García, Rabasco, A., Martínez-Trancón, Rodríguez de Ledesma, I., and Pérez-Regadera. 1993. Characterization of the karyotype of the tench (*Tinca tinca* L.) and analysis of its chromosomal heterochromatic regions by C-banding, Ag-staining, and restriction endonuclease banding. *Cytogenet. Cell Genet.* 62: 220-223.
- Pascale, E., Liu, C., Valle, E., Usdin, K., and Furano, A.V. 1993. The evolution of long interspersed repeated DNA (L1, LINE1) revealed by the analysis of an ancient rodent L1 DNA family. *J. Mol. Evol.* 36: 9-20.
- Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S., and Tanksley, S.D. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127: 181-197.
- Paulsson, G., Höög, C., Bernholm, K., and Wieslander, L. 1992. Balbiani ring 1 gene in *Chironomus tentans*: sequence organization and dynamics of a coding minisatellite. *J. Mol. Biol.* 225: 349-361.
- Pearson, W.R., and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85: 2444-2448.
- Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Meth. Enzym.* 183: 63-98.

Pendás, A.M., Morán, P., and García-Vázquez, E. 1993a. Replication banding pattern in Atlantic salmon (*Salmo salar*). *Genome* 36: 440-444.

Pendás, A.M., Morán, P., and García-Vázquez, E. 1993b. Ribosomal RNA genes are interspersed throughout a heterochromatic chromosome arm in Atlantic salmon. *Cytogenet. Cell Genet.* 63: 128-130.

Pendás, A.M., Morán, P., and García-Vázquez, E. 1993c. Multi-chromosomal location of ribosomal RNA genes and heterochromatin association in brown trout. *Chromosome Res.* 1: 63-67.

Pendás, A.M., Morán, P., and García-Vázquez, E. 1994a. Organization and chromosomal location of the major histone cluster in brown trout, Atlantic salmon and rainbow trout. *Chromosoma* (in press).

Pendás, A.M., Morán, P., and García-Vázquez, E. 1994b. Chromosomal mapping and nucleotide sequence of two tandem repeats of Atlantic salmon 5S rDNA. *Cytogenet. Cell Genet.* 67: 31-36.

Pendás, A.M., Morán, P., and García-Vázquez, E. 1994c. Physical mapping of reiterated gene families in salmonids. poster abstract, 5th Intern. Symp on Genetics in Aquaculture, Dalhousie University, Halifax, Canada, June 19-25, 1994.

Perlman, P.S., Peebles, C.L., and Daniels, C. 1989. *In* Intervening Sequences in Evolution and Development (Stone, E.M., and Schwartz, R.J., eds.). Oxford University press, Oxford, UK.

Perna, N.T., Batzer, M.A., Deininger, P.L., and Stoneking, M. 1992. Alu insertion polymorphism: a new type of marker for human population studies. *Human Biol.* 64: 641-648.

Phillips, R.B. and Ihssen, P.E. 1985. Chromosome banding in salmonid fish: nucleolar organizer regions in *Salmo* and in *Salvelinus*. *Can. J. Genet. Cytol.* 27: 433-440.

Phillips, R.B. and Ihssen, P.E. 1986. Inheritance of Q band chromosomal polymorphisms in lake trout. *J. Hered.* 77: 93-97.

Phillips, R.B. and S.E. Hartley. 1988. Fluorescent banding patterns of the chromosomes of the genus *Salmo*. *Genome* 30: 193-197.

Phillips, R.B. and Pleyte, K.A. 1991. Nuclear DNA and salmonid phylogenetics. *J. Fish Biol.* 39A: 259-275.

- Phillips, R.B., Zajicek, K.D., and Utter, F.M. 1985. Q band chromosomal polymorphisms in chinook salmon, *Oncorhynchus tshawytscha*. Copeia 2: 273-278.
- Phillips, R.B., Pleyte, K.A., and Brown, M.R. 1992. Salmonid phylogeny inferred from ribosomal DNA restriction maps. Can. J. Fish. Aquat. Sci. 49: 2345-2353.
- Pinkel, D., Straume, T., Gray, J.W. 1986. Cytogenetic analysis using quantitative, high sensitive, fluorescence hybridisation. Proc. Natl. Acad. Sci. USA. 83: 2934-2938.
- Pleyte, K.A., Phillips, R.B., and Hartley, S.E. 1989. Q-band chromosomal polymorphisms in arctic char (*Salvelinus alpinus*). Genome 32: 129-133.
- Plohl, M., and Ugarković, D. 1994. Analysis of divergence of *Alphitobius diaperinus* satellite DNA - roles of recombination, replication slippage and gene conversion. Mol. Gen. Genet. 242: 297-304.
- Plucienniczak, A., Skowronski, J., and Jaworski, J. 1982. Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. J. Mol. Biol. 158: 293-304.
- Postlethwait, J.H., Johnson, S.L., Midson, C.N., et al. 1994. A genetic linkage map for the zebrafish. Science 264: 699-703.
- Power, G. 1981. Stock characteristics and catches of Atlantic salmon (*Salmo salar*) in Quebec, and Newfoundland and Labrador in relation to environmental variables. Can. J. Fish. Aquat. Sci. 38: 1601-1611.
- Prasad, S.S., Harris, L.J., Baillie, D.L., and Rose, A.M. 1991. Evolutionarily conserved regions of *Caenorhabditis* transposable elements deduced by sequence comparison. Genome 34: 6-12.
- Prodöhl, P.A., Taggart, J.B., and Ferguson, A. 1994. Cloning of highly variable minisatellite DNA single-locus probes for brown trout (*Salmo trutta* L.) from a phagemid library. In Genetics and Evolution of Aquatic Organisms. (Beaumont, A.R., ed.). Chapman and Hall, London. pp. 263-270.
- Prosser, J., Frommer, M., Paul, C., and Vincent, P.C. 1986. Sequence relationships of three human satellite DNAs. J. Mol. Biol. 187: 145-155.
- Radic, M.Z., Saghbini, M., Elton, T.S., Reeves, R., and Harkalo, B.A. 1992. Hoechst 33258, distamycin A, and high mobility group protein I (HMG-I) compete for binding to mouse satellite DNA. Chromosoma 101: 602-608.

- Radice, A.D., Bugaj, B., Fitch, D.H. and Emmons, S.W. 1994. Ancient origin of the Tc1 transposon family. *Mol. Gen. Genet.* (in press).
- Reed, K.M., Beukeboom, L.W., Eickbush, D.G., and Werren, J.H. 1994. Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. *J. Mol. Evol.* 38: 352-362.
- Rico, C., Kuhnlein, U., and Fitzgerald, G.J. 1991. Spawning patterns in the three-spined stickleback (*Gasterosteus aculeatus* L.): an evaluation by DNA fingerprinting. *J. Fish Biol.* 39A: 151-158.
- Robertson, H. 1993. The *mariner* transposable element is widespread in insects. *Nature* 362: 241-245.
- Robins, D.M., and Samuelson, L.C. 1992. Retrotransposons and the evolution of mammalian gene expression. *Genetica* 86: 191-201.
- Rogaev, E.I. 1990. Simple human DNA-repeats associated with genomic hypervariability, flanking the genomic retroposons and similar to retroviral sites. *Nucl. Acids Res.* 18: 1879-1885.
- Rose, M.R., and Doolittle, W.F. 1983. Molecular biological mechanisms of speciation. *Science* 220: 157-162.
- Rosenzweig, B., Liao, L. and Hirsh, D. 1983. Sequence of the *C. elegans* transposable element Tc1. *Nucl. Acids Res.* 11: 4201-4209.
- Rossi, M.S., Reig, O.A., and Zorzopulos, J. 1990. Evidence for rolling-circle replication in a major satellite DNA from the South American rodents of the genus *Ctenomys*. *Mol. Biol. Evol.* 7: 340-350.
- Royle, N.J., Clarkson, R.E., Wong, Z., and Jeffreys, A.J. 1988. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* 3: 353-360.
- Royle, N.J., Baird, D.M., and Jeffreys, A.J. 1994. A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nature Genet.* 6: 52-56.
- Ruskin, B., and Fink, G.R. 1993. Mutations in POL1 increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. *Genetics* 133: 43-56.
- Ryman, N. 1983. Patterns of distribution of biochemical genetic variation in

salmonids: differences between species. *Aquaculture* 33: 1-21.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual* (2nd ed.). Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY. 3 vols.

Samuelson, L.C., Wiebauer, K., Snow, C.M., and Meisler, M.H. 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell. Biol.* 10: 2513-2520.

Sánchez, L., Abuín, M., and Amaro, R. 1993. Cytogenetic characterization of the AS cell line derived from the Atlantic salmon (*Salmo salar* L.). *Cytogenet. Cell. Genet.* 64: 35-38.

Sánchez, L., Martínez, P., Viñas, A., and Bouza, C. 1990. Analysis of the structure and variability of nucleolar organizer regions of *Salmo trutta* by C-, Ag-, and restriction endonuclease banding. *Cytogenet. Cell Genet.* 54: 6-9.

Sánchez, L., Martínez, P., and Bouza, C., and Viñas, A. 1991. Chromosomal heterochromatin differentiation in *Salmo trutta* with restriction enzymes. *Heredity* 66: 241-249.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74: 5463-5467.

Schartl, M., Erbeling-Denk, Holter, S., Nanda, I., Schmid, M., Schroder, J.H., and Epplen, J.T. 1993. Reproductive failure of dominant males in the peциliid fish *Limia perugiae* determined by DNA fingerprinting. *Proc. Natl. Acad. Sci. USA.* 90: 7064-7068.

Schlötterer, C., and Tautz, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: 211-215.

Schmidt, E.R. 1984. Clustered and interspersed repetitive DNA sequence family of *Chironomus*. The nucleotide sequence of the Cla-elements and of various flanking sequences. *J. Mol. Biol.* 178: 1-15.

Schmidtke, J., Schmitt, E., Matzke, E., and Engel, W. 1979. Non-repetitive DNA sequence divergence in phylogenetically diploid and tetraploid t^hostan species of the family Cyprinidae and the order Isopondyli. *Chromosoma* 75: 185-198.

Schwarzacher-Robinson, Cram, Meyne, and Moyzis. 1988. Characterization of human heterochromatin by *in situ* hybridization with satellite DNA probes. *Cytogenet. Cell*

Genet. 47: 192-196.

Scribner, K.T., Arntzen, J.W., and Burke, T. 1994. Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite and multilocus minisatellite data. *Mol. Biol. Evol.* 11: 737-748.

Shimizu, Y., Yoshida, K., Ren, C.-S., Fujinaga, K., Rajagopalan, S., and Chinnadurai, G. 1983. *Hinf* family: a novel repeated DNA family of the human genome. *Nature* 302: 587-590.

Singer, M.F. 1982. Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol.* 76: 67-112.

Singer, M.F. and J. Skowronski. 1985. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. *TIBS* 10: 119-22.

Singer, M.F., and Berg, P. 1991. *Genes and Genomes, a Changing Perspective*. University Science Books, Mill Valley, California.

Sinnet, D., Deragon, J.-M., Simard, L.R., and Labuda, D. 1990. Alu-morphs - human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers. *Genomics* 7: 331-334.

Siracusa, L.D., Jenkins, N.A., and Copeland, N.G. 1991. Identification and applications of repetitive probes for gene mapping in the mouse. *Genetics* 127: 169-179.

Sletten, A., Olsaker, I., and Lic, O. 1993. Isolation and characterization of (GT)_n repetitive sequences from Atlantic salmon, *Salmo salar* L. *Animal Genet.* 24: 195-197.

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-35.

Smith, G.R., Swirydzuk, K., Kimmel, P.G., and Wilkinson, B.H. 1982. Fish biostratigraphy of late Miocene to Pleistocene sediments of the western Snake River plain, Idaho. *Idaho Bur. Mines Geol. Bull.* 26: 519-541.

Southern, E.M. 1975. Long-range periodicities in mouse satellite DNA. *J. Mol. Biol.* 94: 51-69.

Stahl, G. 1987. Genetic population structure of Atlantic salmon. In *Population Genetics and Fisheries Management*. (Ryman, N., and Utter, F., eds.). University of

Washington Press, Seattle, Washington. pp. 121-140.

Stavenhagen, J.B., and Robins, D.M. 1988. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 55: 247-254.

Stearley, R.F., and Smith, G.R. 1993. Phylogeny of the Pacific trouts and salmon, *Oncorhynchus*, and genera of the family Salmonidae. *Trans. Am. Fish. Soc.* 122: 1-33.

Stephan, W. 1989. Tandem-repetitive noncoding DNA: forms and forces. *Mol. Biol. Evol.* 6: 298-312.

Stephan, W., and Cho, S. 1994. Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* 136: 333-341.

Strachan, T., Webb, D., and Dover, G.A. 1985. Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J.* 4: 1701-1708.

Stuart, G.R., Dixon, B. and Pohajdak, B. 1992. Isolation of a putative retrovirus pol gene fragment from trout. *Comp. Biochem. Physiol.* 102B: 137-142.

Suzuki, H., Miyashita, N., Moriwaki, K., Kominami, R., Muramatsu, R., Kanehisa, M.T., Bonhomme, F., Petras, M.L., Yu, Z. and Lu, D. 1986. Evolutionary implication of heterogeneity of the nontranscribed spacer region of ribosomal DNA repeating units in various subspecies of *Mus musculus*. *Mol. Biol. Evol.* 3: 126-137.

Taggart, J.B., and Ferguson, A. 1990a. Hypervariable minisatellite DNA single locus probes for the Atlantic salmon (*Salmo salar*). *J. Fish Biol.* 37: 991-993.

Taggart, J.B., and Ferguson, A. 1990b. Minisatellite DNA fingerprints of salmonid fishes. *Anim. Genet.* 21: 377-389.

Taggart, J.B., Hynes, R.A., Prodöhl, P.A., and Ferguson, A. 1992. A simplified protocol for routine total DNA isolation from salmonid fishes. *J. Fish. Biol.* 40: 963-965.

Takeaki, I., Abe, S., Yamaha, E., and Yamazaki, F. 1993. A preliminary study of BrdU replication banding in the chromosomes of chum salmon, *Oncorhynchus keta*. *Chromosome Info. Serv.* 54: 30-32.

Tautz, D. 1993. Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In *DNA Fingerprinting: the State of the Science.* (Pena, S.D.J., Chakraborty, R., Epplen, J.T., and Jeffreys, A.J., eds.). Birkhäuser Verlag, Basel. pp. 21-28.

Tautz, D., Trick, M., and Dover, G.A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656.

Thacker, J., Chalk, J., Ganesh, A., and North, P. 1992. A mechanism for deletion formation in DNA by human cell extracts: the involvement of short sequence repeats. *Nucl. Acids Res.* 20: 6183-6188.

Thomas, W.K., and Beckenbach, A.T. 1989. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. *J. Mol. Evol.* 29: 233-245.

Treco, D., and Arnheim, N. 1986. The evolutionarily conserved repetitive sequence d(TGAC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* 6: 3934-3947.

Trinh, T.Q., and Sinden, R.R. 1993. The influence of primary and secondary DNA-structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* 134: 409-422.

Turner, B.J., Elder, J.F., and Laughlin, T.F., and Davis, V. P. 1990. Genetic variation in clonal vertebrates detected by simple-sequence DNA fingerprinting. *Proc. Natl. Acad. Sci. USA.* 87: 5653-5657.

Turner, B.J., Elder, J.F., Jr., and Laughlin, T.F. 1991. Repetitive DNA sequences and the divergence of fish populations: some hopeful beginnings. *J. Fish Biol.* 39A: 131-142.

Unfried, K., Schiebel, K., and Hemleben, V. 1991. Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not in *Vigna angularis*. *Gene* 99: 63-68.

VanderWiel, P.L., Voytas, D.F., and Wendel, J.F. 1993. *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). *J. Mol. Evol.* 36: 429-447.

Varvio, S.-L., and Kaukinen, J. 1993. Bovine microsatellites: racial differences and association with SINE-elements. In *DNA Fingerprinting: the State of the Science*. (Pena, S.D.J., Chakraborty, R., Epplen, J.T., and Jeffreys, A.J., eds.). Birkhäuser Verlag, Basel. pp. 437-443.

Viñas, A., Gómez, C., Martínez, P., and Sánchez, L. 1994. Induction of G-bands on *Anguilla anguilla* chromosomes by the restriction endonucleases *HaeIII*, *HinfI*, and *MseI*. *Cytogenet. Cell Genet.* 65: 79-81.

- Vissel, B., and Choo, K.H.A. 1992. Evolutionary relationships of multiple alpha satellite subfamilies in the centromeres of human chromosomes 13, 14, and 21. *J. Mol. Evol.* 35: 137-146.
- Vogt, P. 1990. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Hum. Genet.* 84: 301-336.
- Vos, J.C., van Luenen, H.G.A.M., and Plasterk, R.H.A. 1993. Characterization of the *Caenorhabditis elegans* Tc1 transposase *in vivo* and *in vitro*. *Genes and Dev.* 7: 1244-1253.
- Wahls, W.P., Wallace, L.J., and Moore, P.D. 1990. The Z-DNA motif d(TG)₃₀ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol. Cell Biol.* 10: 785-793.
- Wakimoto, B., Kalfayan, L.J., and Spradling, A.C. 1986. Developmentally regulated expression of *Drosophila* chorion genes introduced at diverse chromosomal positions. *J. Mol. Biol.* 187: 33-45.
- Walsh, J.B. 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics* 115: 553-67.
- Waring, M., and Britten, R.J. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* 154: 791-794.
- Waye, J.S. and H.F. Willard. 1989. Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc. Natl. Acad. Sci. USA.* 86: 6250-6254.
- Weiner, A.M., Deininger, P.L., and Efstratiadis, A. 1986. Non-viral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Ann. Rev. Biochem.* 55: 631-661.
- Welsh, J., and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18: 7213-7218.
- Wevrick, R., and Willard, H.F. 1989. Long-range organization of tandem arrays of α satellite DNA at the centromeres of human chromosomes: high-frequency array length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. USA.* 86: 9394-9398.
- White, T.J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR protocols: a*

guide to methods and applications. (Innis, M.A., Gelford, D.H., Srinisky, J., and White, T.J., eds.). Academic Press, NY. pp. 315-322.

Wiberg, U.H. 1983. Sex determination in the European eel (*Anguilla anguilla*, L.). Cytogenet Cell Genet. 36: 589-598.

Willard, H.F. 1989. The genomics of long tandem arrays of satellite DNA in the human genome. Genome 31: 737-744.

Willard, H.F., and J.S. Wayne. 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. TIG 3: 192-198.

Williams, S.M., DeSalle, R. and Strobeck, C. 1985. Homogenization of geographical variants at the nontranscribed spacer rDNA in *Drosophila mercatorum*. Mol. Biol. Evol. 2: 338-346.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18: 6531-6535.

Wilson, M.V.H. 1974. Fossil fishes of the Tertiary of British Columbia, Ph.D. thesis, U. of Toronto, Toronto.

Wilson, M.R., Marcuz, A., van Ginkel, F., Miller, N.W., Clem, L.W., Middleton, D., and Warr, G.W. 1990. The immunoglobulin M heavy chain constant region gene of the channel catfish, *Ictalurus punctatus*: An unusual mRNA splice pattern produces the membrane form of the molecule. Nucl. Acids Res. 18: 5227-5233.

Winberg, B.C., Zhou, Z., Dallas, J.F., McIntyre, C.L., and Gustafson, J.P. 1993. Characterization of minisatellite sequences from *Oryza sativa*. Genome 36: 978-983.

Winkfein, R.J., Moir, R.D., Krawetz, S.A., Blanco, J., States, J.C., and Dixon, G.H. 1988. A new family of repetitive, retroposon-like sequences in the genome of the rainbow trout. Eur. J. Biochem. 176: 255-264.

Wirgin, I.I., Grunwald, C., and Garte, S.J. 1991. Use of DNA fingerprinting in the identification and management of a striped bass population in the southeastern United States. Trans. Amer. Fish. Soc. 120: 273-282.

Wright, J.M. 1989. Nucleotide sequence, genomic organization and evolution of a major repetitive DNA family in tilapia (*Oreochromis mossambicus/hornorum*). Nucl. Acids Res. 17: 5071-5079.

- Wright, J.M. 1993. DNA fingerprinting of fishes. *In* Biochemistry and molecular biology of fishes. Vol. 2. (Hochachka, P., and Mommsen, T., eds.). Elsevier, New York. pp. 57-91.
- Wright, J.M. 1994. Mutation at VNTRs: are minisatellites the evolutionary progeny of microsatellites? *Genome* 37: 345-347.
- Yamazaki, F., and Goodier, J. 1993. Cytogenetic effects of hydrostatic pressure treatment to suppress the first cleavage of salmon embryos. *Aquaculture* 110: 51-59.
- Yamazaki, F., Onozato, H., and Arai, K. 1981. The chopping method for obtaining permanent chromosome preparations from embryos of teleost fishes. *Bull. Jpn. Soc. Sci. Fish.* 47: 963.
- Yamazaki, F., Goodier, J., and Yamano, K. 1989. Chromosome aberration caused by aging and hybridization in charr, masu salmon and related salmonids. *Physiol. Ecol. Jpn., Spec.* 1: 529-542.
- Yee, H.A., Wong, A.K.C., van der Sande, J.H., and Rattner, J.B. 1991. Identification of novel single-stranded d(TC)_n binding proteins in several mammalian species. *Nucl. Acids Res.* 19: 949-953.
- Xiong, Y., and Eickbush, T.H. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9: 3353-3362.
- Zhou, L., Sajdak, S.L., and Phillips, R.B. 1994. Minimal intraspecific variation in the sequence of the transcribed spacer regions of the ribosomal DNA of lake trout (*Salvelinus namaycush*). *Genome* 37: 664-671.

